Apr 5th, 12:00 AM

University of Washington eScience Institute: a Data Science Institute Before "Data Science" Was Cool

Micaela Parker
University of Washington eScience Institute

Et al.

UNIVERSITY of WASHINGTON

a Data Science Institute before "data science" was cool

Micaela Parker

and

Sarah Stone

co-Executive Directors

Exponential improvements in technology and algorithms are enabling a revolution in discovery

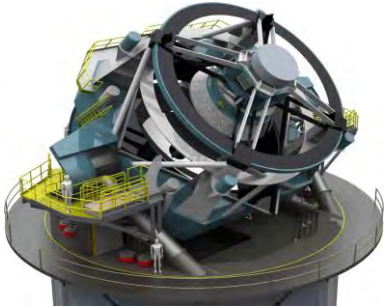
- A proliferation of sensors
- Dramatic instrumentation advancements
- Ever more powerful models producing ever more data
- The creation of almost all information in digital form
- Dramatic cost reductions in storage and computation
- Dramatic improvements in scalability and network bandwidth
- Algorithmic breakthroughs in areas such as machine learning

“Get ready for the fire hose!” – Ginger Armbrust

Data are being collected and used Everywhere!

- Smart homes
- Smart cars
- Smart health
- Smart education
- Smart interaction
  (virtual and augmented reality)
- Smart cities
- Smart discovery **

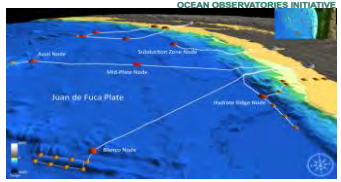
Nearly every field of discovery is transitioning from "data poor" to "data rich"



Astronomy: LSST

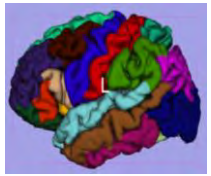
Physics: LHC

Digital Humanities



Oceanography: OOI

Health

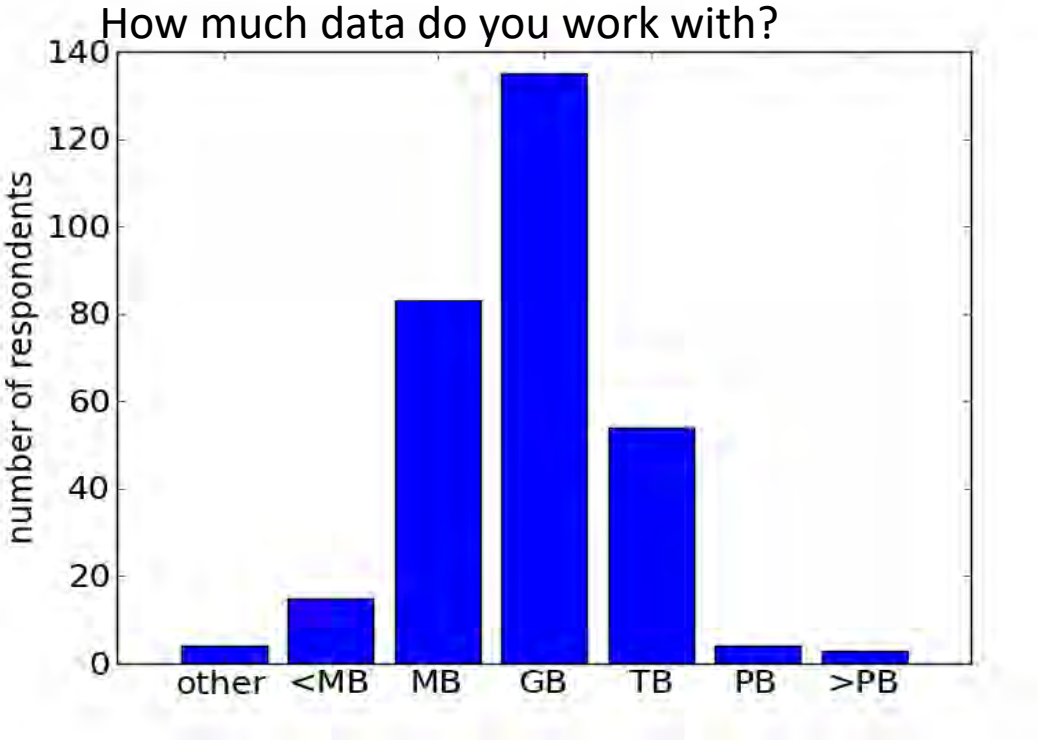


Biology: Sequencing

Economics:
POS terminals

Sociology: Social Media and the Web

Data Science challenges go beyond size



Wright 2013

Inception of the UW eScience Institute

"All across our campus, ***the process of discovery will increasingly rely on researchers' ability to extract knowledge from vast amounts [varieties and velocities] of data****...*

In order to remain at the forefront, UW must be a leader in advancing these techniques and technologies, and in making [them] accessible to researchers in the broadest imaginable range of fields."

Inception of the eScience Institute

Mark Emmert

LSU

2004

UNIVERSITY OF WASHINGTON 1861 LVX SIT

Ed Lazowska, CSE

Tom Daniel, Biology

Werner Stuetzle, Statistics

"From data to knowledge to action"

- The ability to extract knowledge from large, heterogeneous, noisy datasets lies at the heart of 21st century discovery

- To remain at the forefront, researchers *in all fields* will need access to state-of-the-art data science methodologies and tools

- These methodologies and tools will need to advance rapidly, driven by the requirements of discovery

Data science is driven more by ***intellectual infrastructure*** (human capital) and *software infrastructure* (shared tools and services) than by hardware

Major sources of support for our "core effort"

- University of Washington (2008)
  - state line-item for some admin staff and faculty support
- National Science Foundation (2013)
  - 5 years for graduate program development and Ph.D. student funding (IGERT)
- Gordon and Betty Moore Foundation and Alfred P. Sloan Foundation (2013)
  - 5 year partnership with NYU and Berkeley
- Washington Research Foundation (2014)
  - 5 years for faculty recruiting packages, postdocs, new space

UNIVERSITY of WASHINGTON

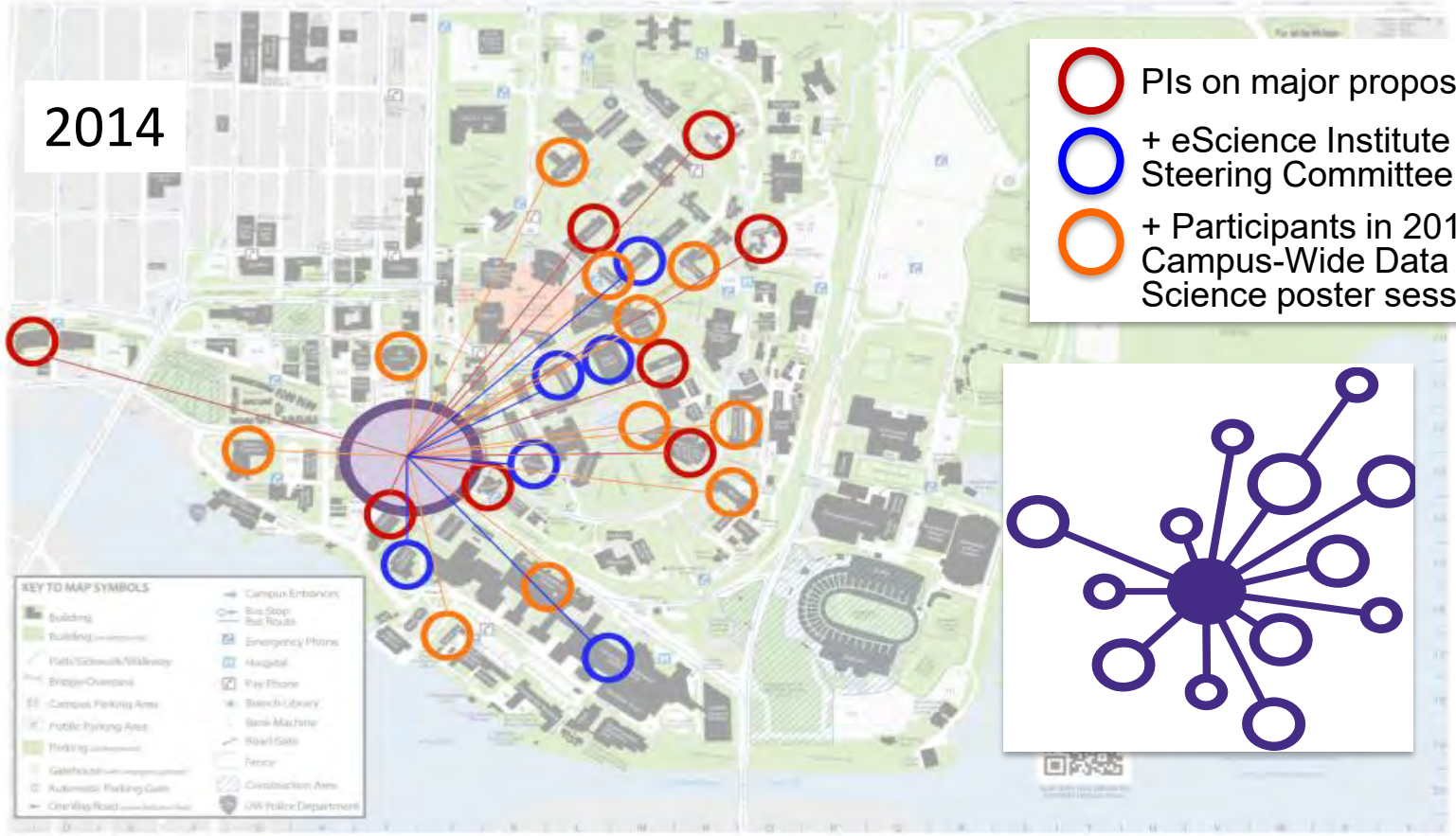
NSF

GORDON AND BETTY MOORE FOUNDATION

ALFRED P. SLOAN FOUNDATION 1934

Washington Research FOUNDATION

Data science: The rising tide that lifts all boats



Our Mission

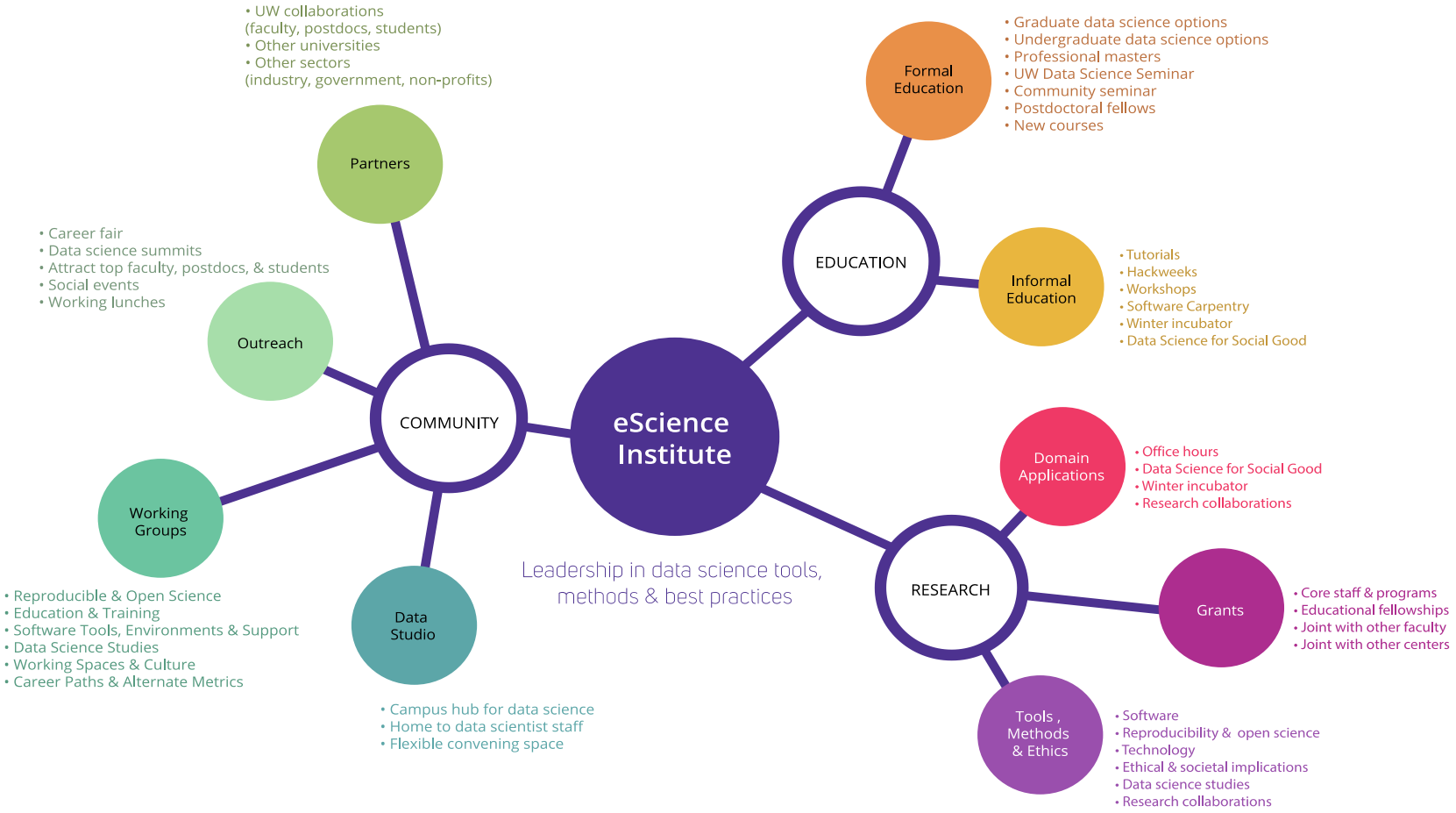
The eScience Institute **empowers** researchers and students in all fields to answer fundamental questions through the use of large, complex, and/or noisy data.

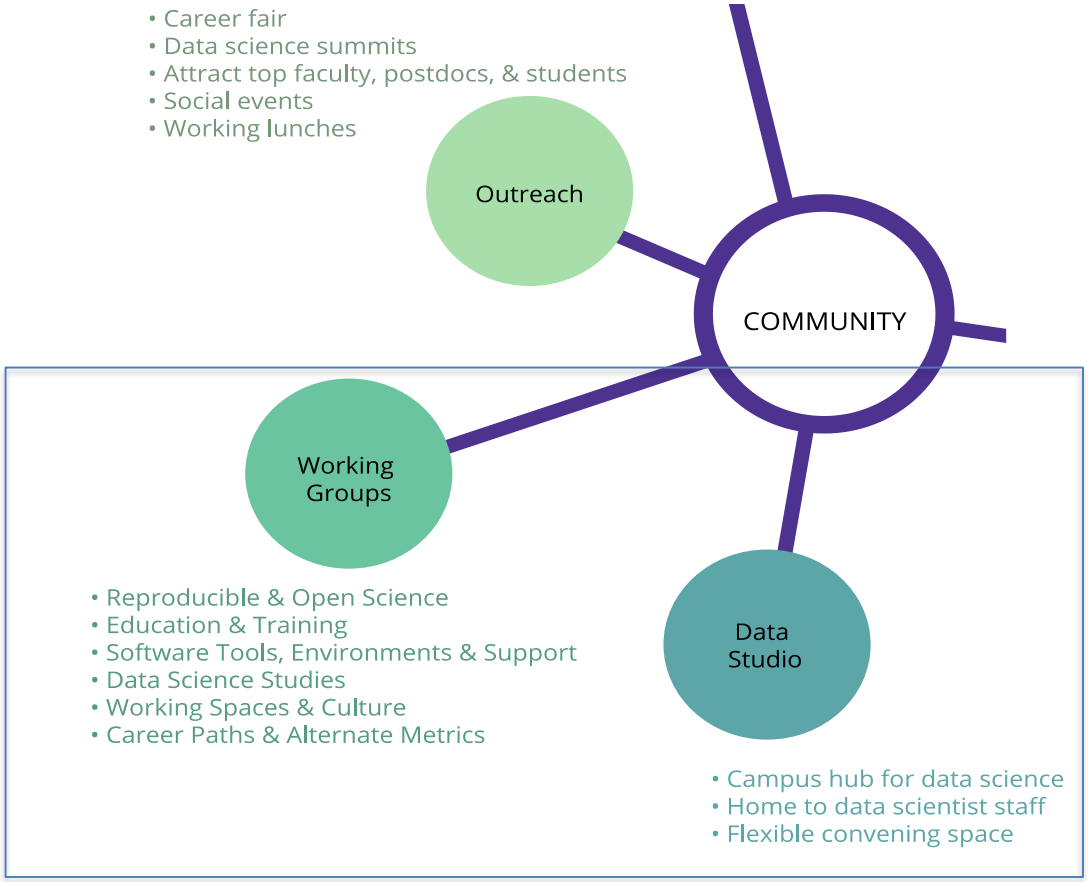
As the **hub** of data-intensive discovery on campus, we lead a **community** of innovators in the techniques, technologies, and best practices of data science and the fields that depend on them.

UNIVERSITY of WASHINGTON
eScience Institute
ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS

Key Activities

- Build a community
- Support data science career paths
- Develop data science education
- Enable scalable research impact
- Advance the state of the art in data science
- Disseminate data science expertise & best practices


• UW collaborations
(faculty, postdocs, students)
• Other universities
• Other sectors
(industry, government, non-profits)
Partners
• Career fair
• Data science summits
• Attract top faculty, postdocs, & students
• Social events
• Working lunches
Outreach
COMMUNITY
Working Groups
• Reproducible & Open Science
• Education & Training
• Software Tools, Environments & Support
• Data Science Studies
• Working Spaces & Culture
• Career Paths & Alternate Metrics
Data Studio
• Campus hub for data science
• Home to data scientist staff
• Flexible convening space
eScience Institute
Leadership in data science tools, methods & best practices
EDUCATION
Formal Education
• Graduate data science options
• Undergraduate data science options
• Professional masters
• UW Data Science Seminar
• Community seminar
• Postdoctoral fellows
• New courses
Informal Education
• Tutorials
• Hackweeks
• Workshops
• Software Carpentry
• Winter incubator
• Data Science for Social Good
RESEARCH
Domain Applications
• Office hours
• Data Science for Social Good
• Winter incubator
• Research collaborations
Grants
• Core staff & programs
• Educational fellowships
• Joint with other faculty
• Joint with other centers
Tools , Methods & Ethics
• Software
• Reproducibility & open science
• Technology
• Ethical & societal implications
• Data science studies
• Research collaborations

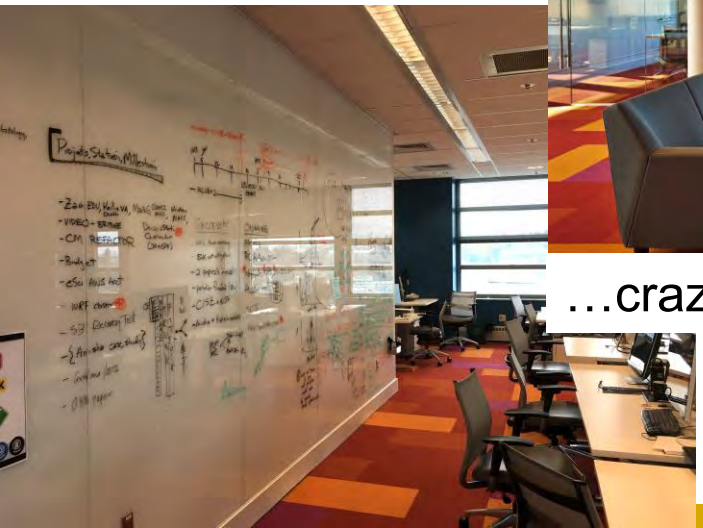

• Career fair
• Data science summits
• Attract top faculty, postdocs, & students
• Social events
• Working lunches
Outreach
COMMUNITY
Working Groups
• Reproducible & Open Science
• Education & Training
• Software Tools, Environments & Support
• Data Science Studies
• Working Spaces & Culture
• Career Paths & Alternate Metrics
Data Studio
• Campus hub for data science
• Home to data scientist staff
• Flexible convening space

WRF Data Science Studio
a campus-wide collaboration space

Build a Community

- Partner with campus libraries
- Take advantage of the “water cooler effect”
- Neutral place on campus for collaboration
- Flexible and transformable spaces
  - Drop-in open workspace, small & large meeting rooms

…crazy carpet and great views



Micaela Parker
Program Coordinator

Chris Mentzel
Gordon and Betty Moore Foundation

Josh Greenberg
Alfred P. Sloan Foundation

The Challenge

University
Domain
Research

Data
Science
Practice

as **data increases in all forms and in all fields**, even some of the very best researchers struggle to generate knowledge and insight from these data

The Grand Experiment

University
Domain
Research

Spur new methods development

Data
Science
Practice

Enable data-driven discovery

Community Building For Insitutions

Working Groups

Career Paths and Alternative Metrics

Working Group Lead: Ed Lazowska

Identify "data science fellows" who might otherwise slip through the cracks or go to industry, and groom them for a new breed of faculty position by creating new roles for data science professionals on campus that are not subject to "publish or perish," and that are equipped to pursue pragmatic, high-impact software-oriented data science projects.

LEARN MORE ABOUT CAREER PATHS AND ALTERNATIVE METRICS >

Reproducibility and Open Science

Working Group Lead: David Beck & Ariel Rokem

Establish a culture of reproducibility and open science, and develop tools to support an environment where researchers find both the tools and the best practices for their research to be openly accessible to society and fully reproducible, more effectively feeding a productive cycle of research.

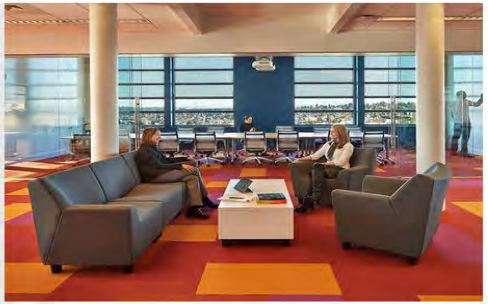
LEARN MORE ABOUT REPRODUCIBILITY AND OPEN SCIENCE >

Working Spaces and Culture

Working Group Leads: Micaela Parker, Sarah Stone

Establish new physical spaces on our campuses, specifically designed to meet the new requirements of data science activities, which in many cases will flourish best outside of traditional departmental boundaries.

LEARN MORE ABOUT WORKING SPACES AND CULTURE >

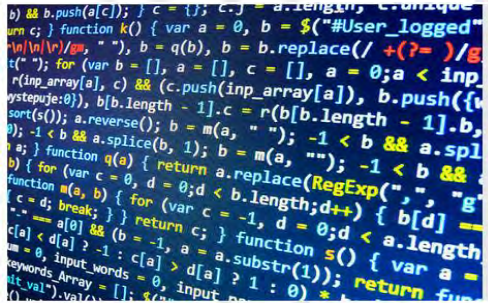


Education and Training

Working Group Lead: Magda Balazinska

Establish alternative mechanisms that are free from departmental politics and conventional structures: boot camps, summer schools, tutorials.

LEARN MORE ABOUT EDUCATION AND TRAINING >



Software Tools, Environments, and Support

Working Group Lead: Jake Vanderplas

Successful projects are characterized by a balance between specialization and generality: sufficiently focused to actually solve a problem, but with the ability to scale to enough users or enough domains to amortize the cost of the initial development. We seek to institutionalize these patterns of success to help deliver the "next 100 Sloan Digital Sky Surveys."

LEARN MORE ABOUT SOFTWARE TOOLS, ENVIRONMENTS, AND SUPPORT >

Data Science Studies

Working Group Co-Lead: Cecilia Aragon
Working Group Co-Lead: Brittany Fiore-Gartland

UW Data Science Studies is a group of cross-disciplinary researchers studying the sociocultural and organizational processes around the emerging practice of data science.

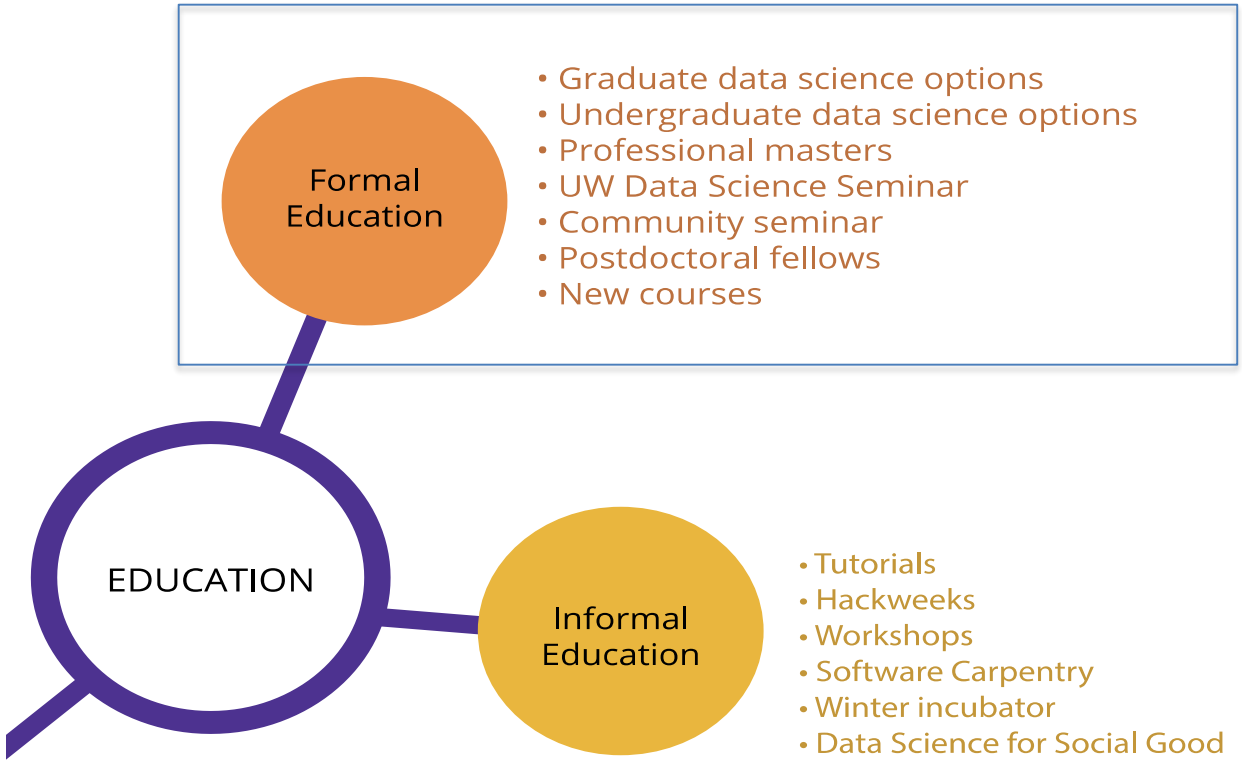
LEARN MORE ABOUT DATA SCIENCE STUDIES >

Reproducibility and Open Science Working Group

- Monthly seminars, meetings and national workshops at UW (2014), Berkeley (2015), NYU (2016)
- Draft guidelines for reproducible research: http://uwescience.github.io/reproducible/
- Case Studies book:

Kitzes, J., Turek, D., & Deniz, F. (Eds.). (2018). The Practice of Reproducible Research: Case Studies and Lessons from the Data-Intensive Sciences. Oakland, CA: University of California Press.

- Tutorials on Tools for reproducibility and open science
  - E.g. GitHub, KnitR, iPython Notebook
- Prototype reproducibility peer review workflow with badges
- Reproducibility roadshows


Formal Education
• Graduate data science options
• Undergraduate data science options
• Professional masters
• UW Data Science Seminar
• Community seminar
• Postdoctoral fellows
• New courses
EDUCATION
Informal Education
• Tutorials
• Hackweeks
• Workshops
• Software Carpentry
• Winter incubator
• Data Science for Social Good

Educational objectives

Goal 1: Educate ALL students in data science

- Introductory courses with no pre-requisites
- Tracks to advanced course work within departmental majors

Goal 2: Create interdisciplinary communities of data science practitioners

Goal 3: Provide both formal curricular training pathways as well as informal “ad hoc” training opportunities

Data Science Education at UW

Under leadership of eScience Education Working Group

- Introductory Courses for pre-Majors
  - Data & Society (Sociology), Introduction to Data Science (CSE, iSchool, and stats), Data Programming in Python (CSE)

- Undergraduate and Graduate Transcriptable Options
  - Two levels: Data Science Option (for tool users)
    Advanced Data Science Option (for tool builders)

Approach => Departments offer data science transcriptable options following developed framework, tailored to needs of major

Broadening access: data science education infrastructure

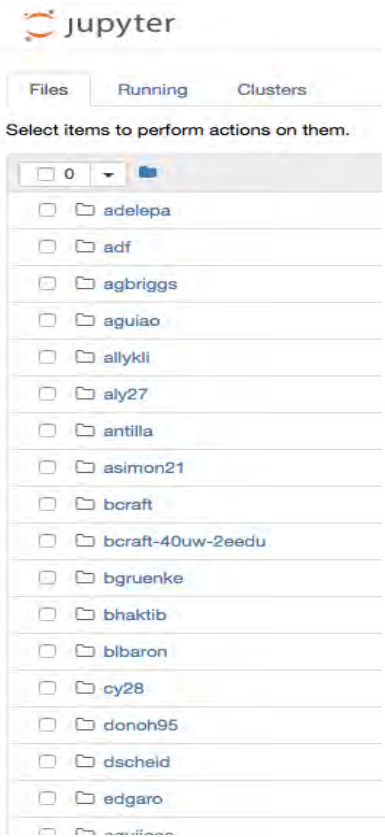
jupyterhub

Multi-user version of Jupyter Notebooks: great for classrooms!

- ✓ **UC Berkeley Foundations of Data Science (Data 8) course with 1,000+ students – the fastest growing class in campus history**
- ✓ **2017 Extended infrastructure beyond Berkeley campus**
  - **UW Human-Centered Data Science (DATA 512)**
  - **UW Information Visualization (HCDE 411)**

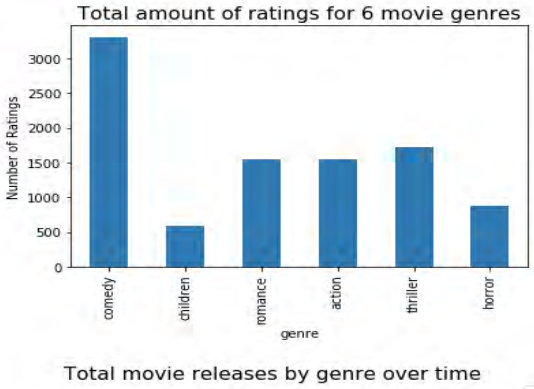
MSDSE
Moore - Sloan
Data Science Environments

JupyterHub in the classroom

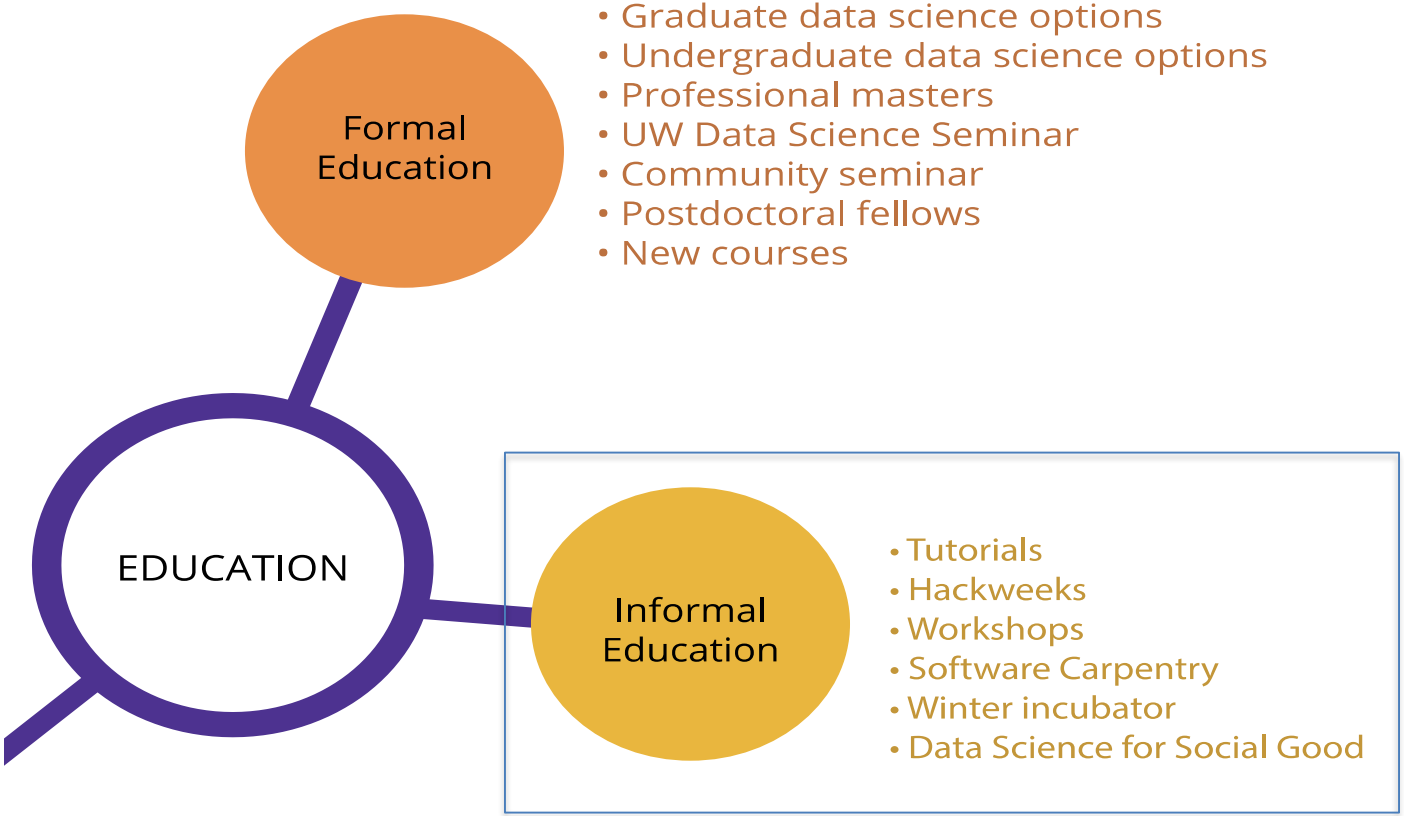



```
# Uses pandas and matplotlib to plot the THRILLER time series plot which is the bottom part of the
# second visualization (small multiples) to show the amount of releases the genre has had from the
# year 1920 to the year 2020.
totaltags4 = ['year', 'Thriller']
df4 = pd.read_table('thrillerovertime.csv', sep= ',', names= totaltags4)
df4.plot(legend = True, x = 'year')
axes = plt.gca()
axes.set_ylim([0,120])
axes.spines['bottom'].set_color('#000000')      # Similar nine lines of code as the previous section.
axes.spines['top'].set_color('#ffffff')
axes.spines['right'].set_color('#ffffff')
axes.spines['left'].set_color('#ffffff')
axes.tick_params(axis='y', colors='white')
plt.gca().get_lines()[0].set_color("red")
plt.gca().get_legend().legendHandles[0].set_color('red')
my_xticks = axes.get_xticks()
plt.xticks([my_xticks[0], my_xticks[-1]], visible=True, rotation="horizontal")
plt.legend(bbox_to_anchor=(1.05, 1), loc=2, borderaxespad=0.)
```

Out[26]: <matplotlib.legend.Legend at 0x7fdbb4496e48>



Data and photos from HCDE 411, courtesy of Brock Craft


Formal Education
• Graduate data science options
• Undergraduate data science options
• Professional masters
• UW Data Science Seminar
• Community seminar
• Postdoctoral fellows
• New courses
EDUCATION
Informal Education
• Tutorials
• Hackweeks
• Workshops
• Software Carpentry
• Winter incubator
• Data Science for Social Good

UNIVERSITY of WASHINGTON
eScience Institute
ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS

Research Consulting Staff

Data Scientists (fully supported)

Jake VanderPlas
Director of Open Software
Ph.D., Astronomy

Ariel Rokem
Sr Data Scientist
Ph.D., Neuroscience

Valentina Staneva
Data Scientist
Ph.D., Applied Math&Stats

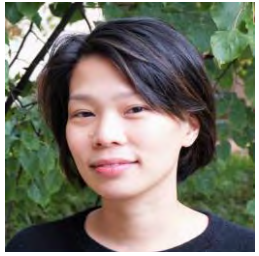
Bernease Herman
Data Scientist
BS, Statistics
was SE at Amazon

Research Faculty

David Beck
Director of Research,
Ph.D. Medicinal Chemistry, Biomolecular Structure & Design

Research IT

Rob Fatland
Director of Cloud and Data Solutions
Senior Data Science Fellow
Ph.D. Geophysics



Amanda Tan
Cloud Technology Lead
Ph.D., Civil and Environmental Engineering

Research Scientists (partial support)

Bryna Hazelton
Research Scientist
Ph.D., Astrophysics

Vaughn Iverson
Sr Research Sci.
PhD,Oceanography

Anthony Arendt
Research Scientist
Ph.D., Geophysics, APL

Joe Hellerstein
Sr Data Sci. Fellow
IBM Research, MSR, Google (ret.)

Jose Hernandez
Research Scientist
Ph.D., Statistics, CCER

Data Science Ethnography Team

Brittany Fiore-Gartland
Director of Ethnography
Ph.D., Communication

Anissa Tanweer
Ethnographer
Ph.D. Candidate, Communication

- Ethnography is integrated into a wide range of Data Science Environment activities
  - Help us remain reflective and reflexive; what is working, what's not, what can we do better?
  - Help us keep data ethics in every conversation
- Develop ethnography research questions
  - E.g., who does data science, how are they networked, forms of social interaction and organization, intellectual groupings, career reward structures, collaborative tool use in scientific workflows, data science values and ethics, etc.

Community Learning Within Domains

Hackweeks

- domain-focused communities

- week-long, 3 main components:

  - tutorials in state-of-the-art methodology
  - project work in a collaborative environment
  - peer-teaching and -learning

Community Learning Within Domains

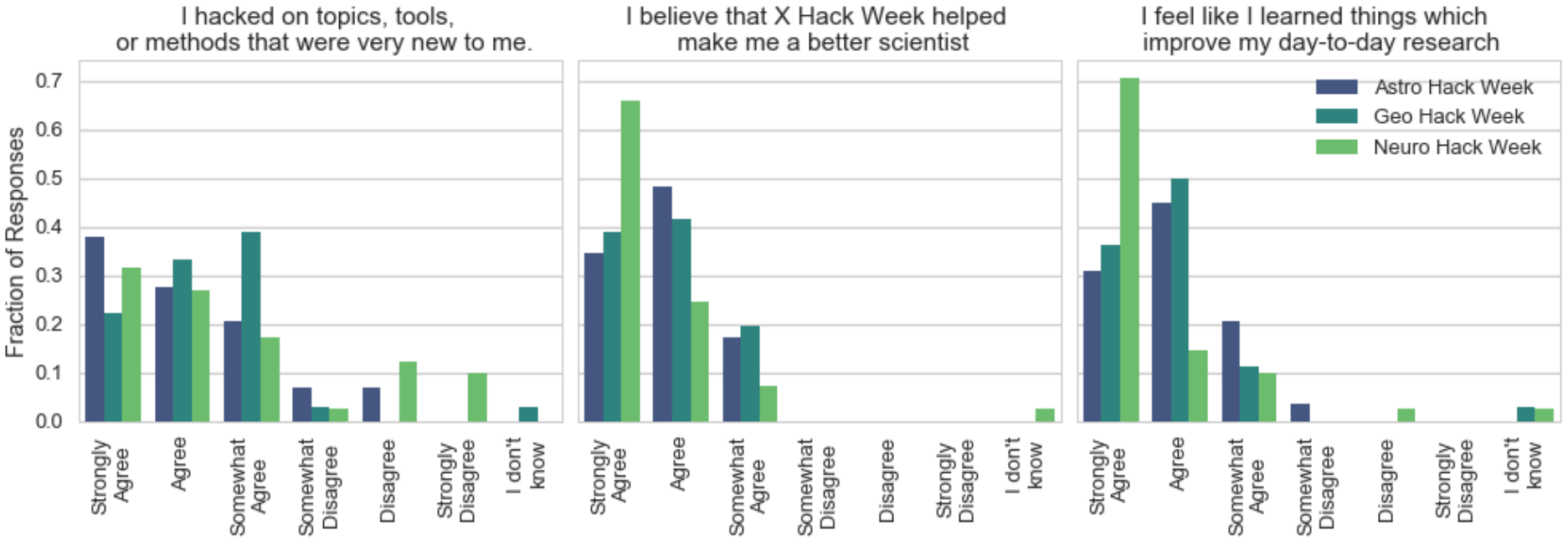
Hackweeks

- domain-focused communities
- week-long, three components:
  - tutorials
  - project work
  - peer-learning

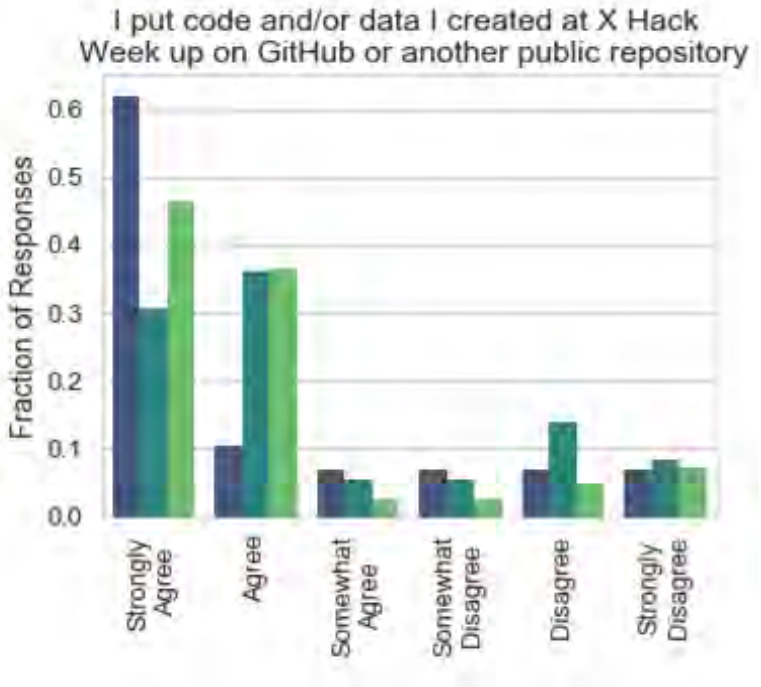
Signs of Success

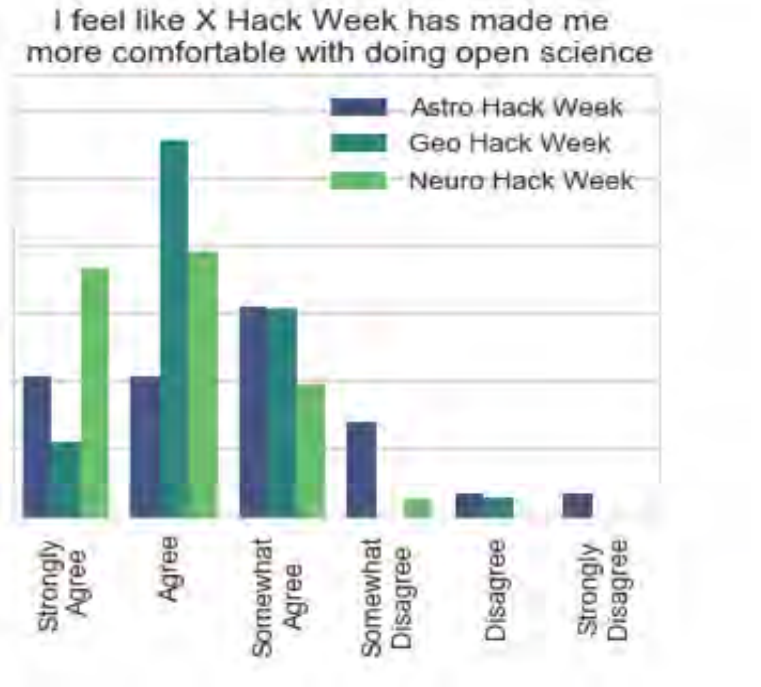
- AstroHackweek's 5th iteration goes international to Leiden, Netherlands in 2018
- NeuroHackweek → NeuroHackademy, 2 week summer program
- New Hackweeks coming online: WaterHack, OceanHack, SocioHack(?)
- Outcomes include papers (8 so far), software (e.g. Stingray) and results (e.g. renewable energy sourcing in the state of WA)

Exit Survey Responses: Research Methods

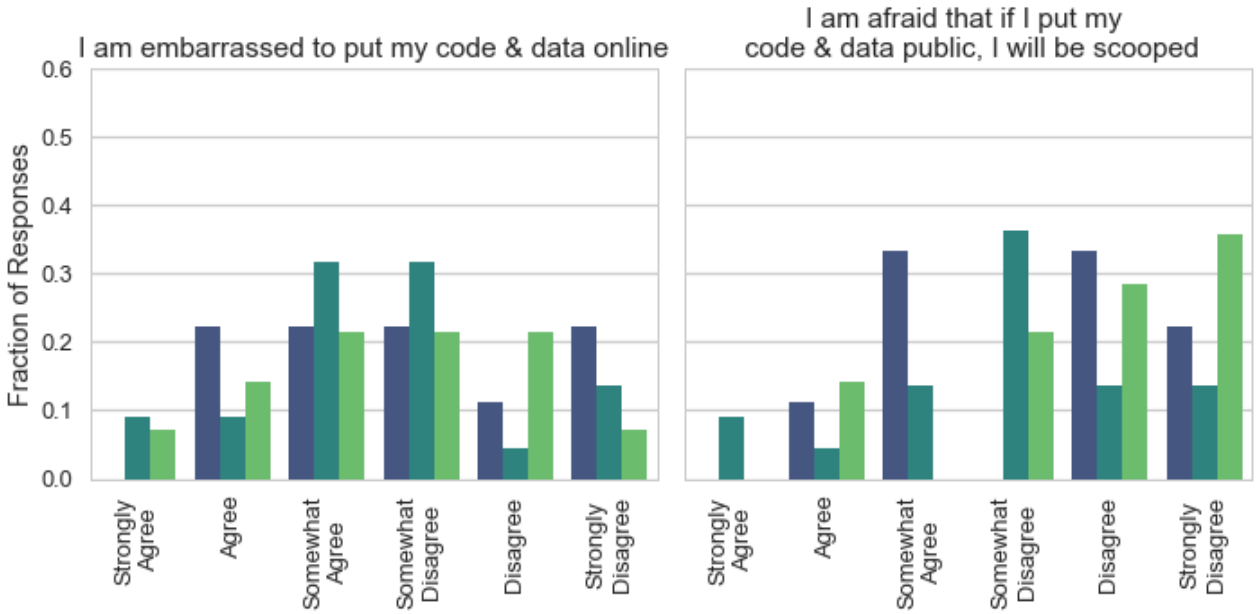


Exit Survey Responses: Open Science





Exit Survey Responses: Open Science



Community Learning Around Methods

XD Working Groups & Workshops

XD's are new methods-focused communities

- host seminars, blogs
- workshops: 2-3 days, include tutorials, talks by experts, and make sessions

Inaugural ImageXD (2016):

- 50 researchers, 14 institutions
- computer vision, microscopy, materials imaging, photography, earth science, neuroscience, astronomy, software development, and more.

Community Learning Around Methods

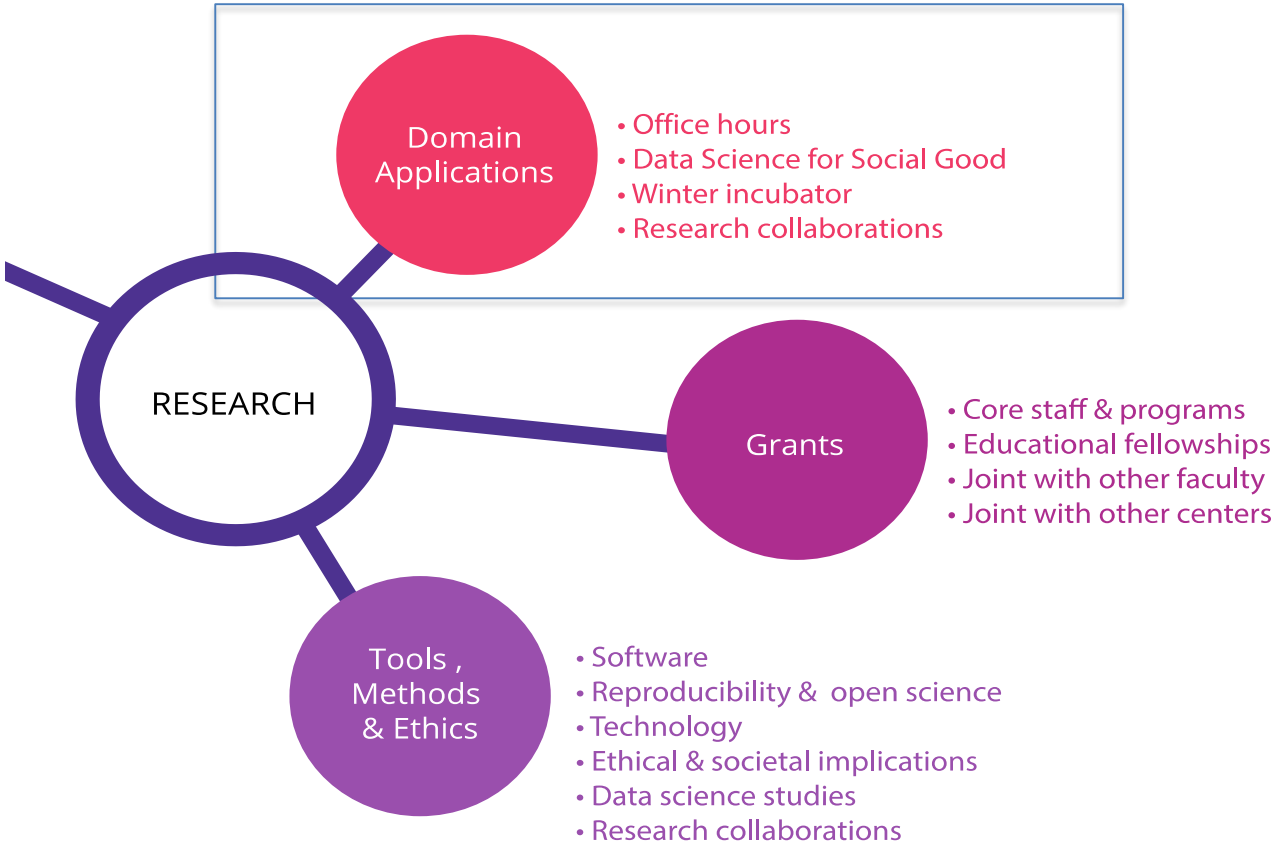
XD Workshops

XD's are new methods-focused communities

- host seminars, blogs
- workshops: 2-3 days, include tutorials, talks by experts, and make sessions

Signs of Success

- ImageXD is on its 3rd iteration, spawned: TextXD (in 2017), GraphXD (in 2018)
- Example outcomes: blueprints for open source image processing, training sets for ML applications, analysis projects


Domain Applications
• Office hours
• Data Science for Social Good
• Winter incubator
• Research collaborations
RESEARCH
Grants
• Core staff & programs
• Educational fellowships
• Joint with other faculty
• Joint with other centers
Tools , Methods & Ethics
• Software
• Reproducibility & open science
• Technology
• Ethical & societal implications
• Data science studies
• Research collaborations

Office Hours

http://escience.washington.edu/office-hours/

UNIVERSITY of WASHINGTON eScience Institute
ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS

UW-IT

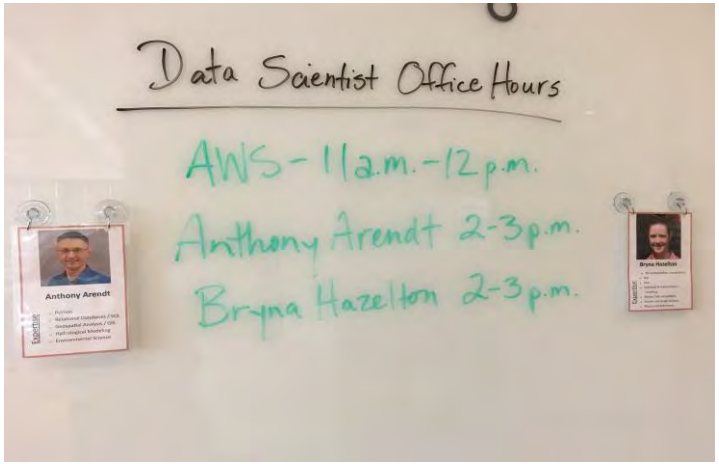
amazon web services

CENTER for STATISTICS and the SOCIAL SCIENCES

+ a b | e a u

UNIVERSITY LIBRARIES

Google Cloud Platform



Scalable Research Impact

(The space between Office Hours and Grant Proposals)

Data Science Incubator Programs

- Move from “accidental” encounters to engineered partnerships
- Identify emerging opportunities around campus
- Provide a shared environment where researchers can learn from an in-house team, external mentors, *and each other*

Winter Incubator Program

- Quarter-long, in-Studio projects, engagement *two days per week*
  - Project Lead + Data Scientist

the "ah ha" moment!

- 4-6 concurrent projects: Network effects among cohort beyond 1:1 interactions
- 1-2 page proposals describing a shovel-ready problem, the science, and how a solution can generalize to other groups
- Participation from faculty, grad students, staff

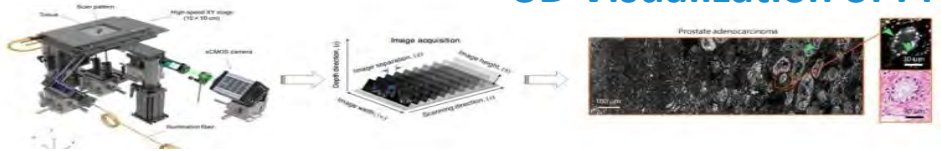
What we're looking for: Projects where fruitful collaboration is possible, with potential for significant impact. "Do only what we can only do together"

Incubator Projects

- No Limits!
- Domain backgrounds in Astronomy, Neuroscience, Applied Math, Physics, Statistics, Earth Science, Oceanography, Finance
- Strong expertise in acoustic and image processing and analytics, machine learning, Python, cloud computing, and big data systems
- Some recent interest in urban science and civic tech (DSSG)
- “extracting knowledge from large, heterogeneous, or noisy datasets”

NiPy

Examples from 4 years of Incubator

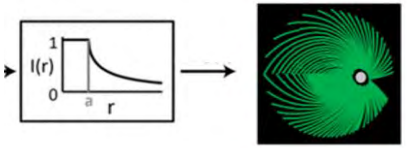


3D Visualization of Prostate Cancer Using Light-Sheet Microscopy

Project Lead: Dr. Nicholas Reder, Pathology
eScience Data Scientist: Ariel Rokem

Analysis of Kenya's Routine Health Information System data

Project Leads: Gregoire Lurton, Abraham Flaxman, Emmanuela Gakidou, IHME
eScience Data Scientist: Dan Halperin

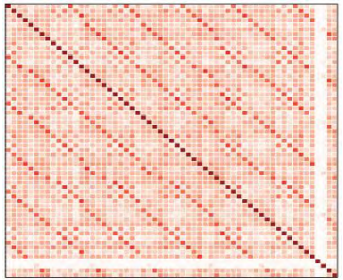


Improved Stimulation Protocols for Sight Restoration Technologies

Project Leads: Ione Fine, Geoffrey M. Boynton, UW Psychology
eScience Data Scientist: Ariel Rokem

Using social media data to identify geographic clustering of anti-vaccination sentiments

Project Leads: Benjamin Brooks, Abraham Flaxman, IHME
eScience Data Scientist: Andrew Whitaker



Methods for Characterizing Human Centromeres

Project Lead: Siva Kasinathan, UW School of Medicine
eScience Data Scientists: Andrew Fiore-Gartland, Bryna Hazelton

Students' sleep and academic performance

Project Lead: Angela Katsuyama, Biology; eScience Data Scientists: Bill Howe and Dan Halperin

UNIVERSITY of WASHINGTON
eScience Institute
DATA SCIENCE FOR SOCIAL GOOD

2017 DSSG Student Fellows

This program brings together students and researchers with data science and domain expertise to work on focused, collaborative projects for societal benefit.

Project Teams

4 projects supported each summer

Each team consists of:

Project Leads (1-2)

eScience Data Scientist Leads (1-2)

DSSG Student Fellows (4) ← highly competitive!

Stakeholders

Examples from 3 years of DSSG

Open Sidewalk Graph for Accessible Trip Planning
The Taskar Center for Accessible Technology

Predictors of Permanent Housing for Homeless Families
Bill and Melinda Gates Foundation

Mining Online Data for Early Identification of Unsafe Food Products
Institute for Health Metrics and Evaluation, Department of Global Health

Use of ORCA data for improved transit system planning and operation
Washington State Transportation Center

Strengthening capacities, knowledge and data sharing platforms for sustainable development
Conservation International, Vital Signs

Can traffic sensor data detect vehicle cruising?
Seattle Department of Transportation

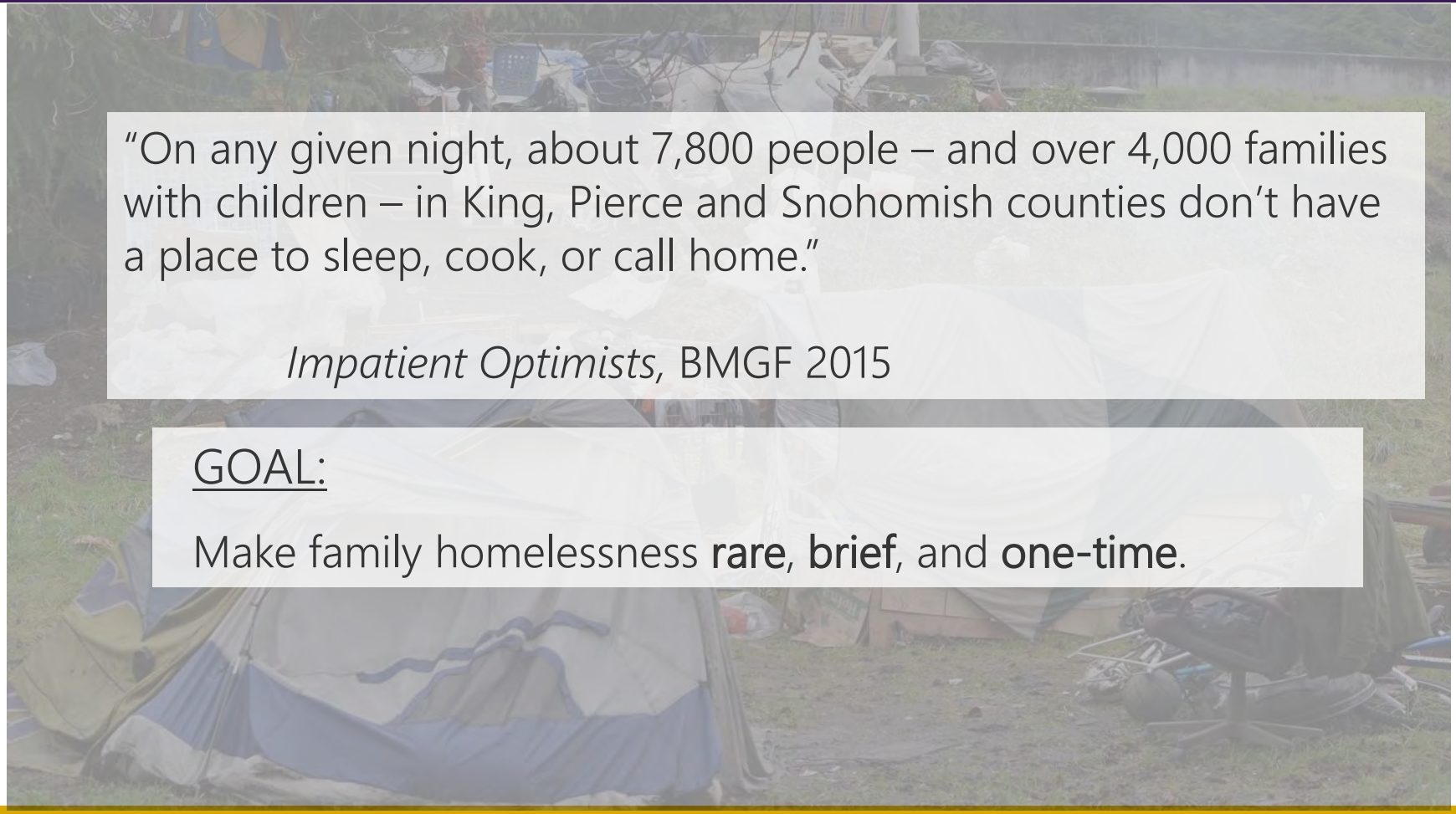
The 'Equity Modeler': examining just development in Seattle
Department of Urban Design and Planning and Department of Architecture

Predictors of Permanent Housing for Homeless Families

BUILDING CHANGES

UNIVERSITY of WASHINGTON eScience Institute
ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS

BILL & MELINDA GATES foundation



"On any given night, about 7,800 people – and over 4,000 families with children – in King, Pierce and Snohomish counties don't have a place to sleep, cook, or call home."

Impatient Optimists, BMGF 2015

GOAL:

Make family homelessness **rare**, **brief**, and **one-time**.

Solomon's Story

Then:

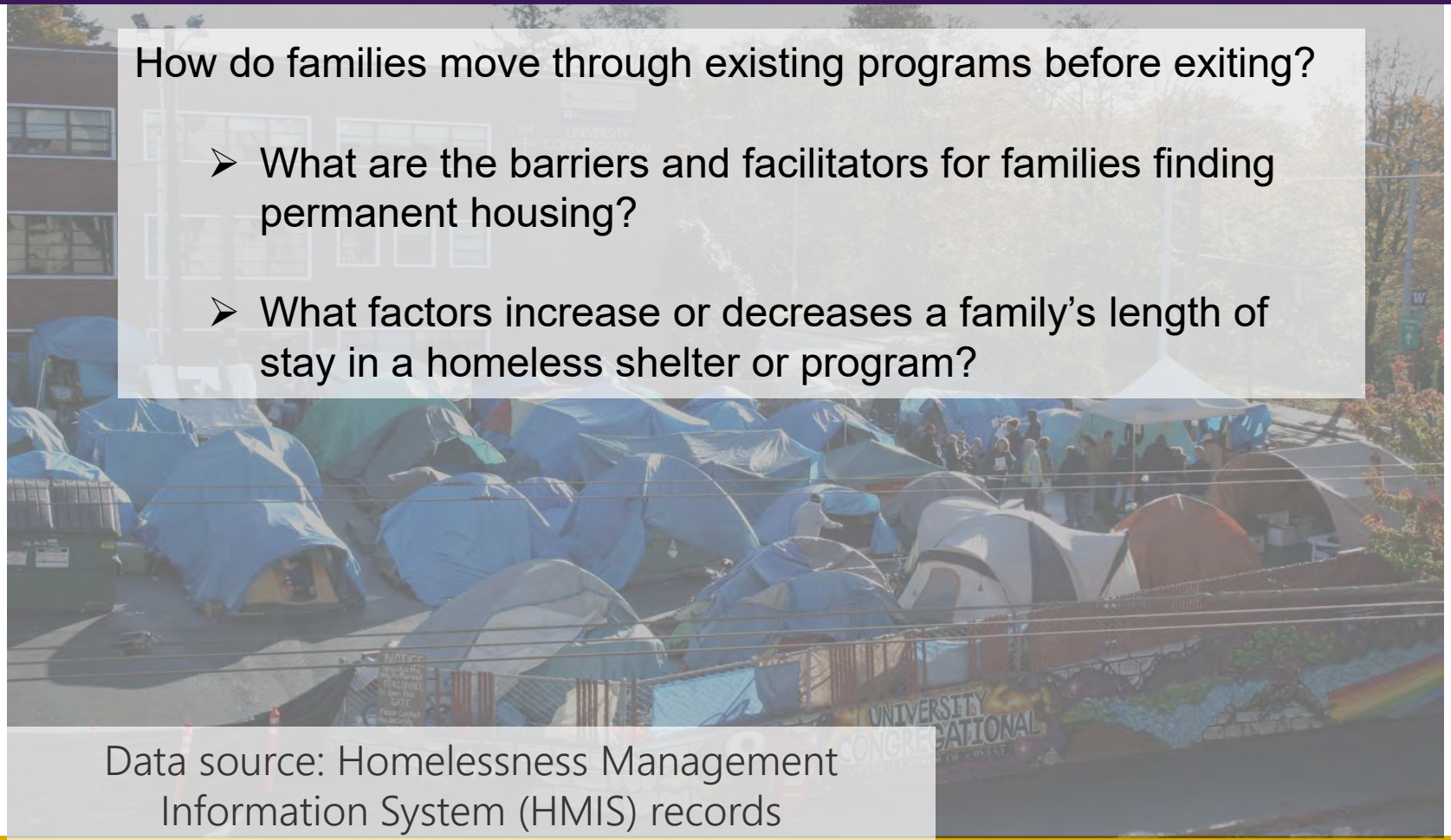
- Refugees to the U.S. (14 years old)
- Family friend housed them but could no longer afford it
- Police took them to Mary's Place
- 8 months in transitional housing: "where you live before you recover and can pay rent".
- Mary's Place helped Solomon land an internship at the Museum of Flight
- He thanks the volunteers at the Mary's Place for helping him apply to college and successfully exit homelessness.

Now:

- Bioengineering student at UW
- volunteers at Mary's Place
- tutors young kids
- cooks meals for the homeless

MARY'S PLACE

"I told myself 'the future is better'"

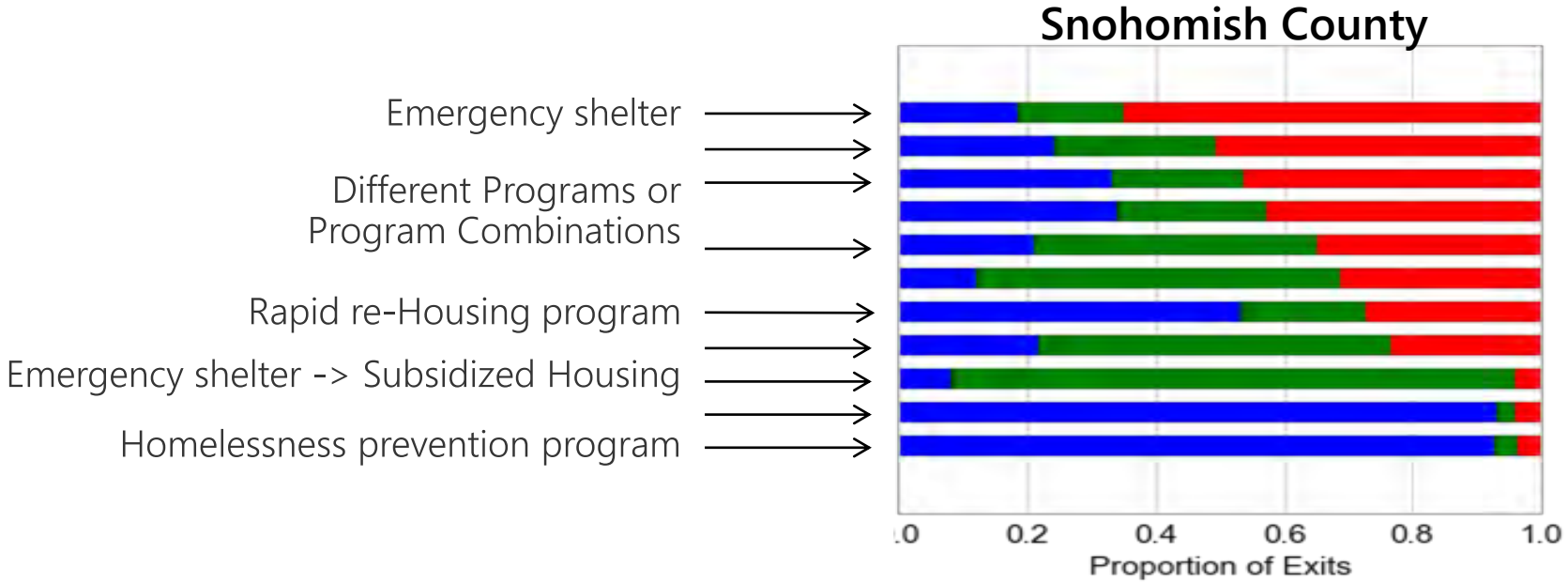


How do families move through existing programs before exiting?

- What are the barriers and facilitators for families finding permanent housing?
- What factors increase or decreases a family's length of stay in a homeless shelter or program?

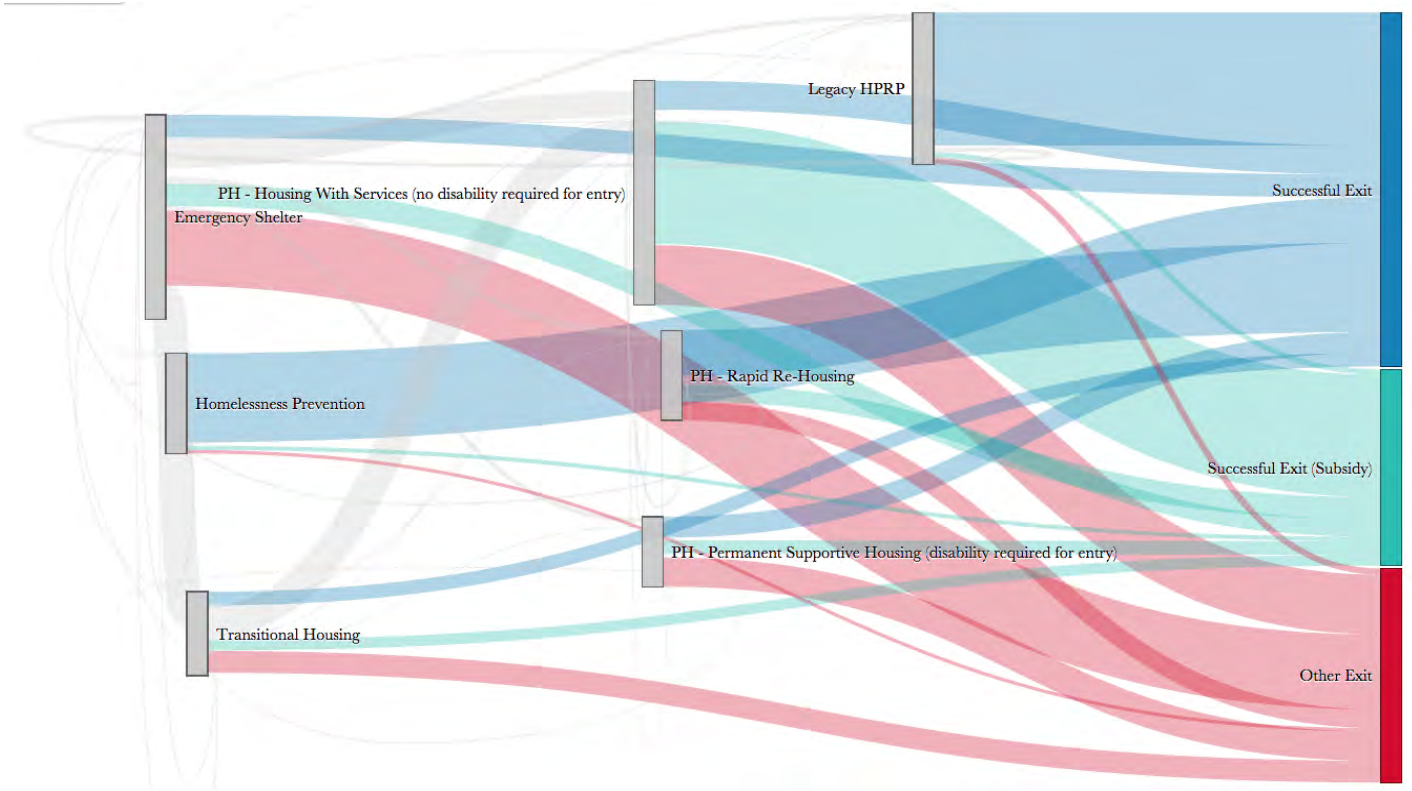
Data source: Homelessness Management Information System (HMIS) records

What program(s) lead to a successful exit?



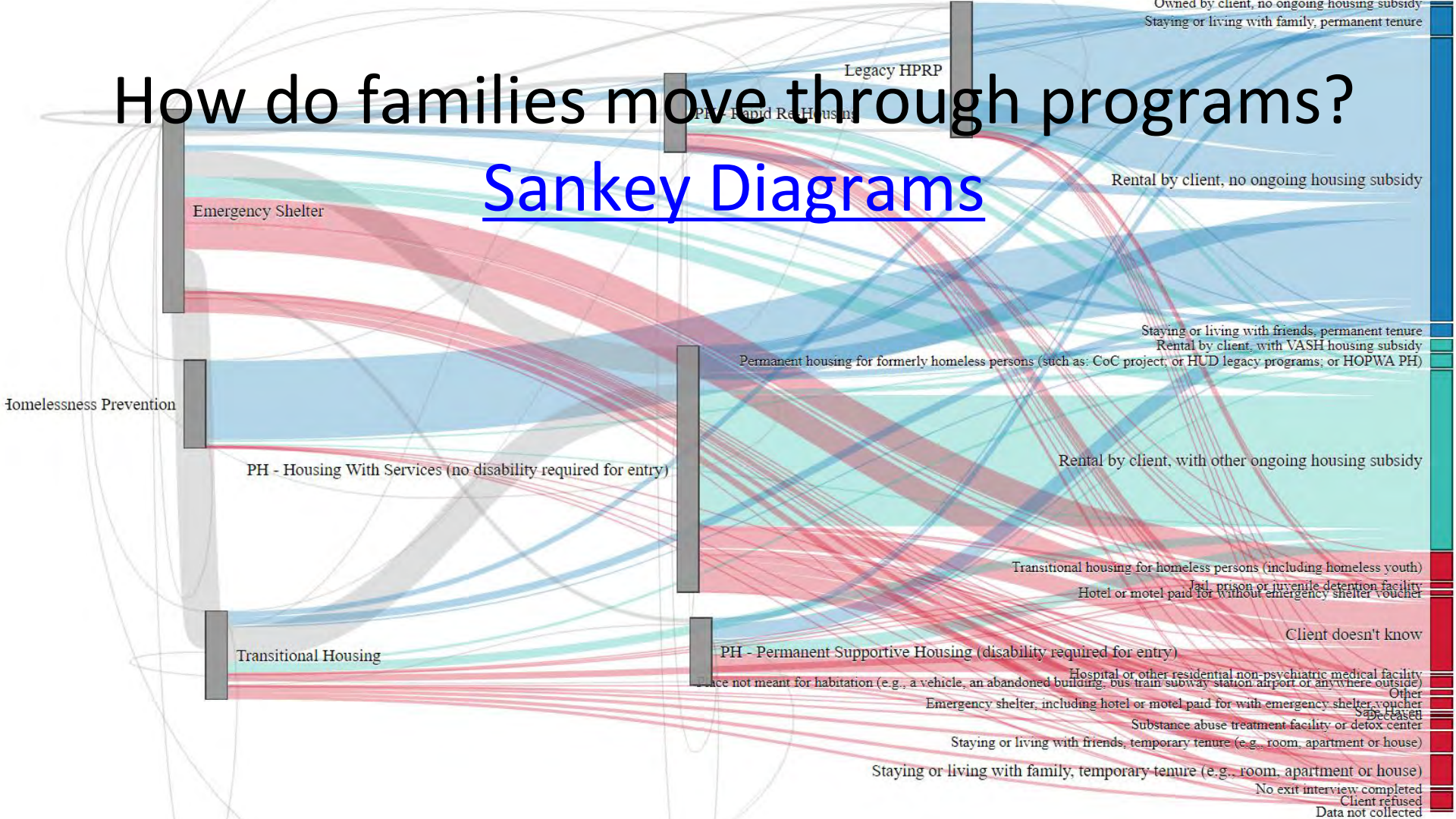
- Successful exit: family found a permanent housing solution
- A proportion of these still receive government subsidies
- Other exits are back into homelessness or to other destinations

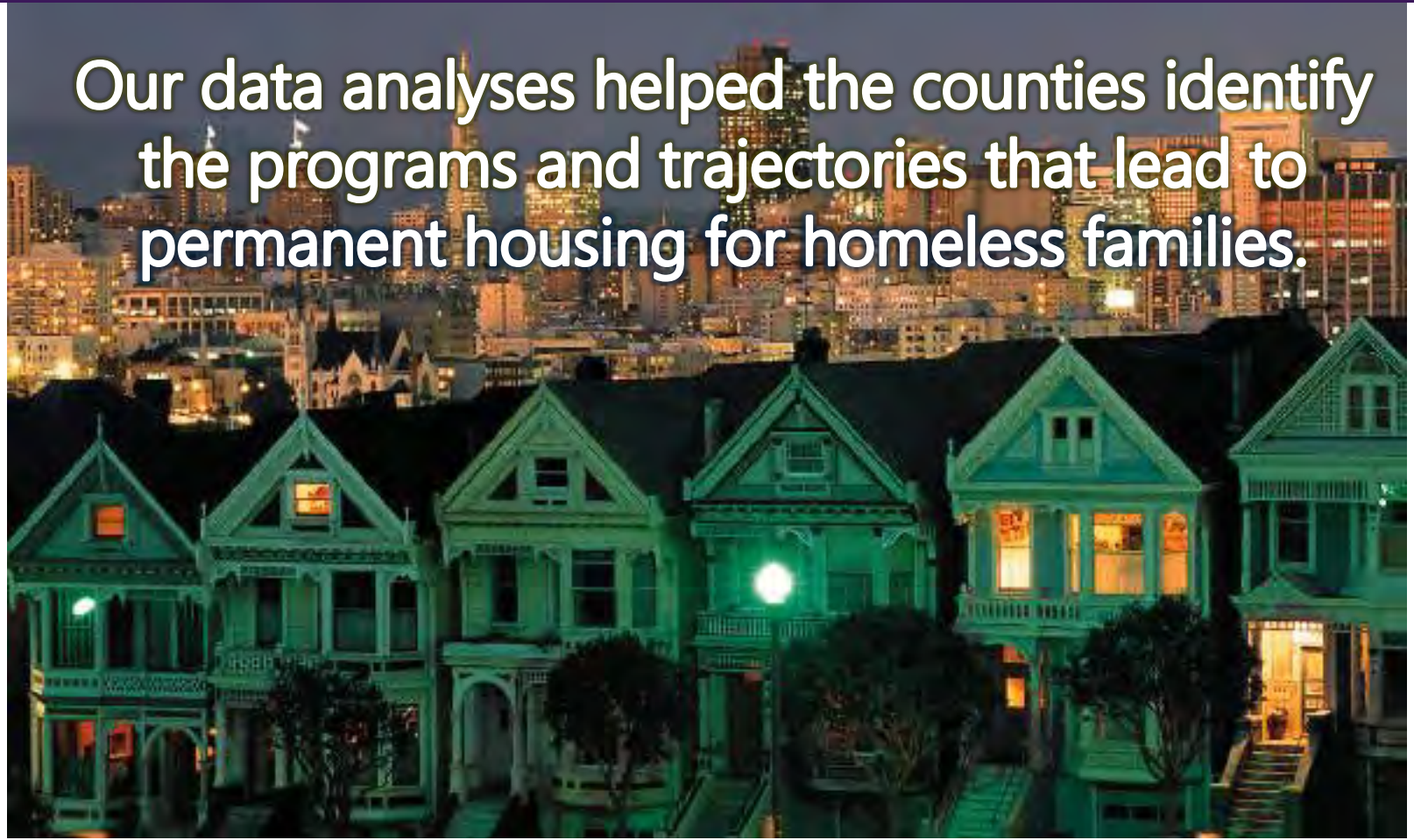
How do families move through programs? : Sankey Diagrams



How do families move through programs?

Sankey Diagrams




Our data analyses helped the counties identify the programs and trajectories that lead to permanent housing for homeless families.


Cities can be incredibly complex to navigate.
For many people, technology provides the information needed to get around.

54.5 million

People in the USA need assistive devices or have trouble walking more than a quarter mile.

U.S. Census Bureau, *Americans With Disabilities: 2010, issued July 2012*

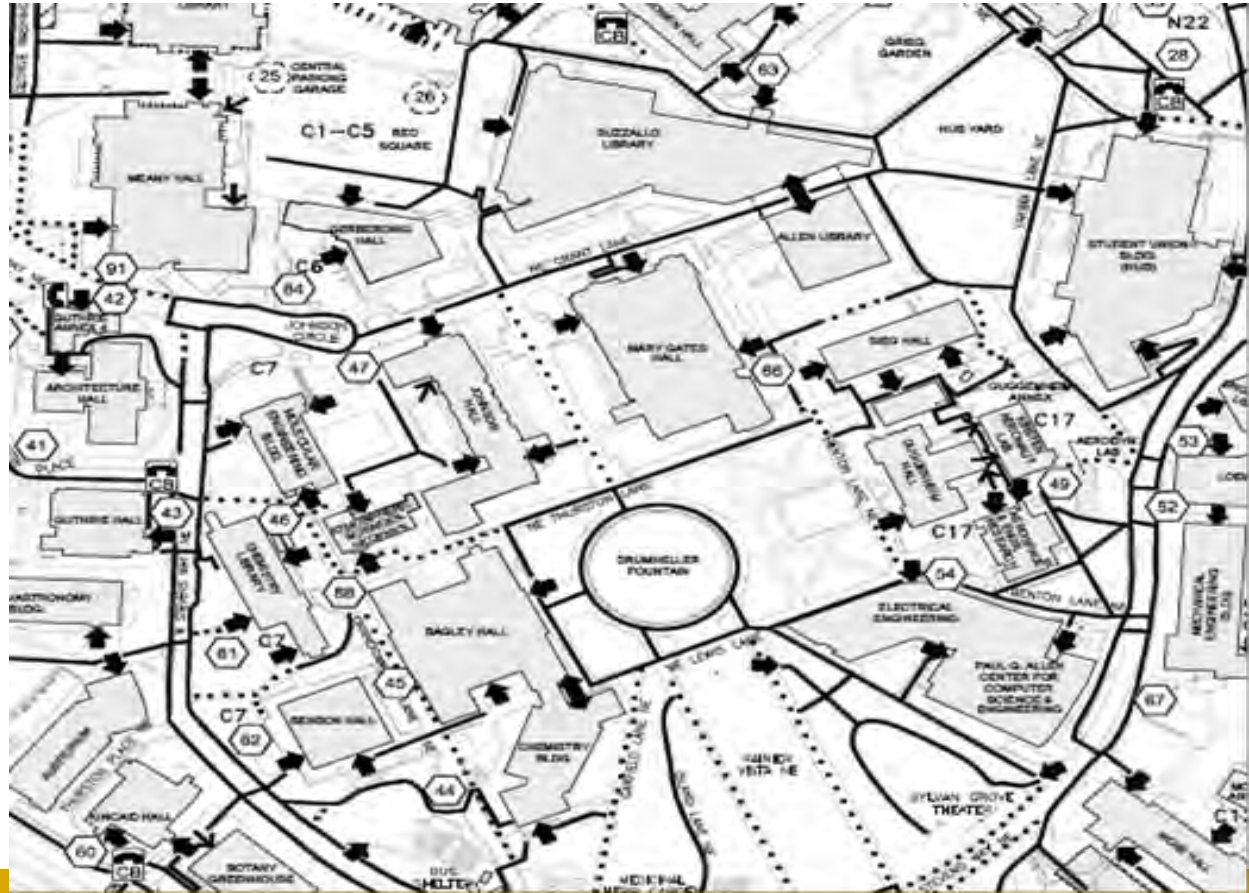
Kevin’s Story

“Using a tool like directions on Google Maps doesn’t really help me get around. Actually sometimes this does more harm than good. I’m sent down streets I can’t cross, or up inclines that are impossible to climb. It can be deeply frustrating.”

Challenge of Current Pedestrian Resources

Static maps:

- Cluttered
- Complex
- Out of date
- Non-editable
- Non-routable



AccessMap

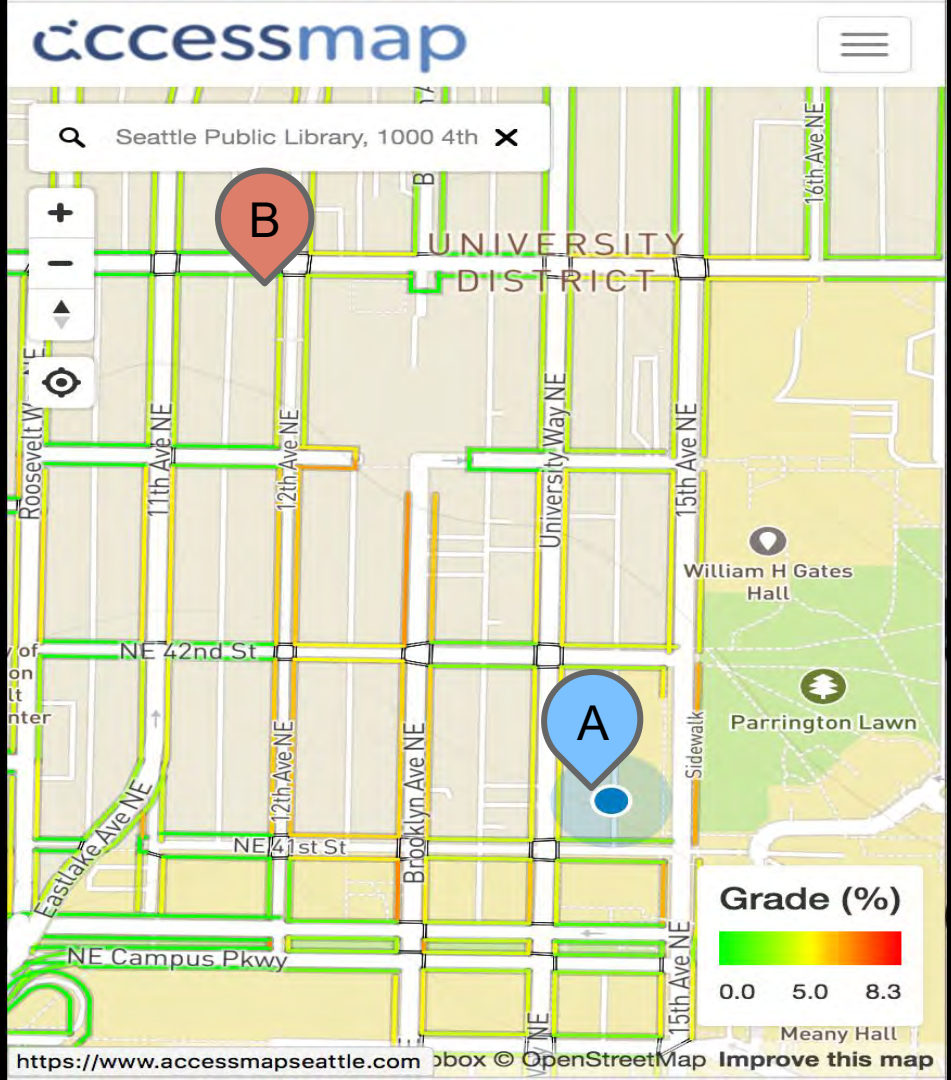
Seattle

TCAT The Taskar Center for Accessible Technology

Trip Planning

I’m here, I want to get there.

What’s the best way?



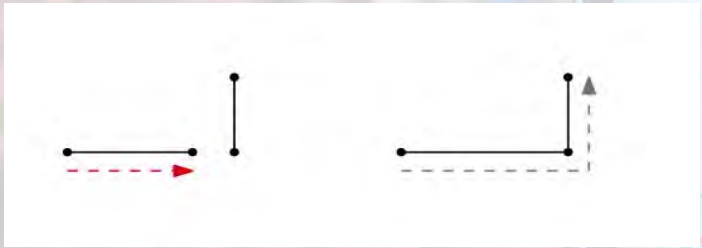
AccessMap

Seattle

TCAT The Taskar Center for Accessible Technology

GOAL: Add data for routing

Connect sidewalks



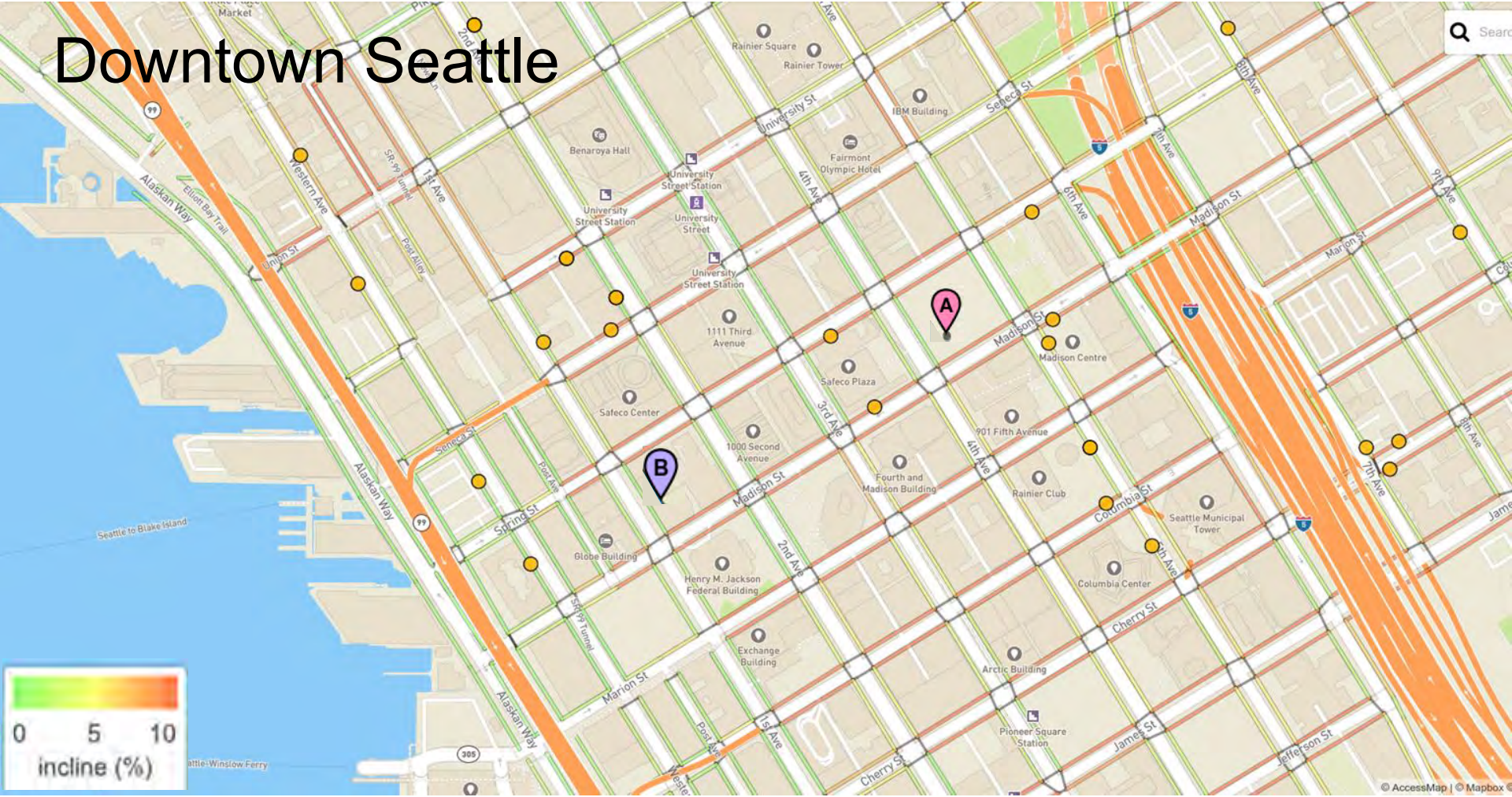
Find the best route

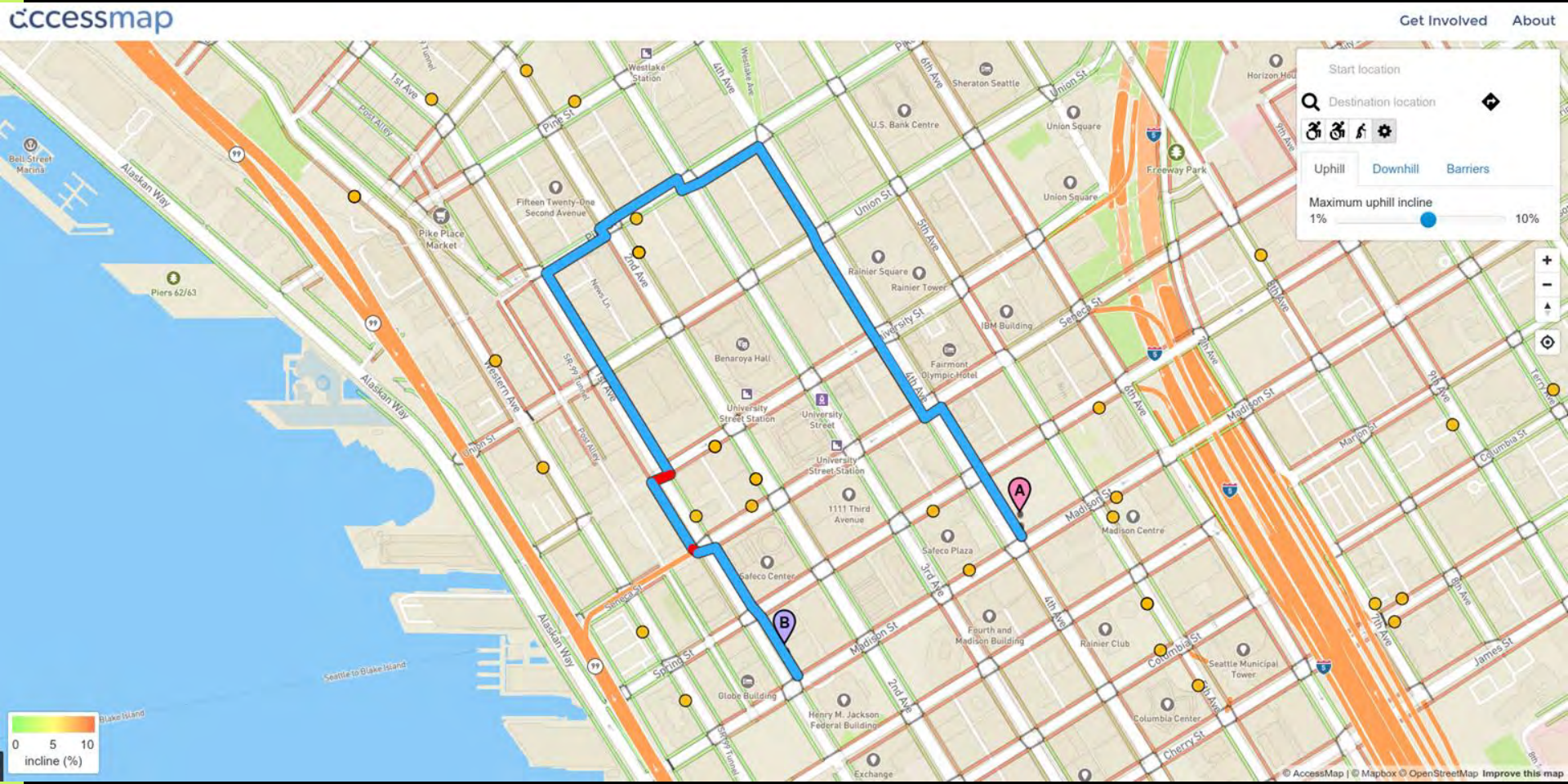
Incline: USGS + Sidewalk lines

Construction: data.seattle.gov Permits

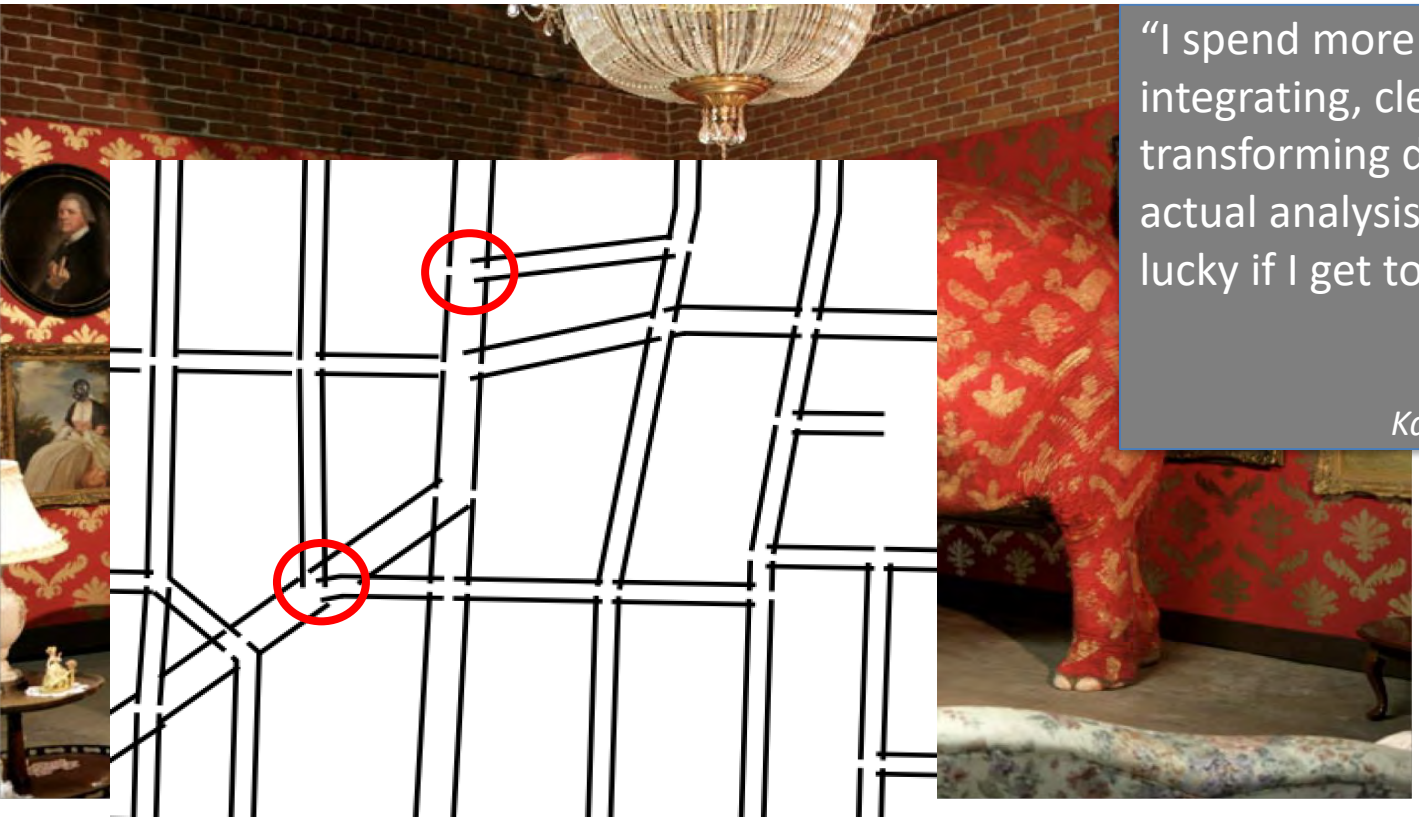
Curb ramps and crosswalks: data.seattle.gov

Downtown Seattle




ccessmap
Get Involved
About
Start location
Destination location
Uphill
Downhill
Barriers
Maximum uphill incline
1%
10%
A
B
0
5
10
incline (%)
Alaskan Way
Pike Place Market
Piers 62/63
Bell Street Marina
Westlake Station
Sheraton Seattle
U.S. Bank Centre
Union Square
Freeway Park
Fifteen Twenty-One Second Avenue
Rainier Square
Rainier Tower
IBM Building
Benaroya Hall
Fairmont Olympic Hotel
University Street Station
1111 Third Avenue
Safeco Plaza
Safeco Center
Madison Centre
Fourth and Madison Building
Rainier Club
Seattle Municipal Tower
Columbia Center
Globe Building
Henry M. Jackson Federal Building
Exchange
Seattle to Blake Island
Blake Island
Union St
Pine St
Madison St
Seneca St
Spring St
Columbia St
James St
Cherry St
Marion St
1st Ave
2nd Ave
3rd Ave
4th Ave
5th Ave
6th Ave
7th Ave
8th Ave
9th Ave
Western Ave
Post Alley
SR-99 Tunnel
© AccessMap | © Mapbox © OpenStreetMap Improve this map

The Reality of Data Science



"I spend more than half of my time integrating, cleansing and transforming data without doing any actual analysis. Most of the time I'm lucky if I get to do any 'analysis' at all."

Anonymous Data Scientist
Kandel et al. 2012 interview study

What I didn't talk about: Data, Responsibly

The Seattle Times

Education | Education Lab | Local News | Transportation

UW student project taps ORCA cards, unlocks data trove

Originally published August 19, 2016 at 10:21 pm | Updated August 21, 2016 at 6:27 am

Trending: Hewlett Packard Enterprise cuts staff as it struggles to keep pace in the cloud

Could Amazon reviews keep you from getting sick? Researchers analyze text to predict food recalls

Trending: Microsoft reveals the 'Xbox Onesie' and the internet goes nuts

Could data help solve Seattle's transportation challenges?

BY CLARE MCGRANE on August 20, 2016 at 3:30 pm

Budding UW Data Scientists Use Their Powers for Social Good

Benjamin Romano
August 24th, 2015

@bromano @xconomy Like Us

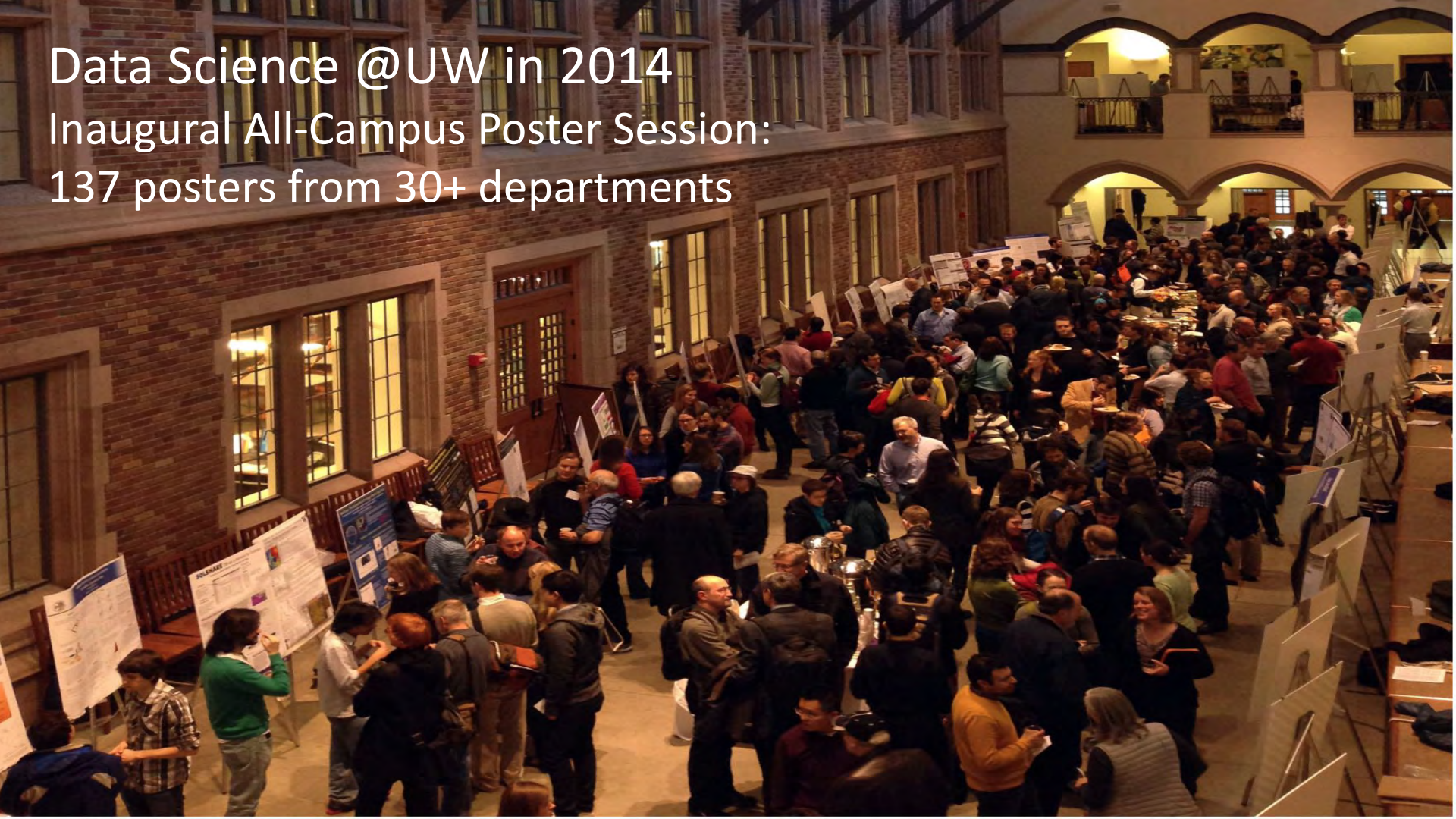
Student projects leapfrog governments and industry in 'Data Science for Social Good' program

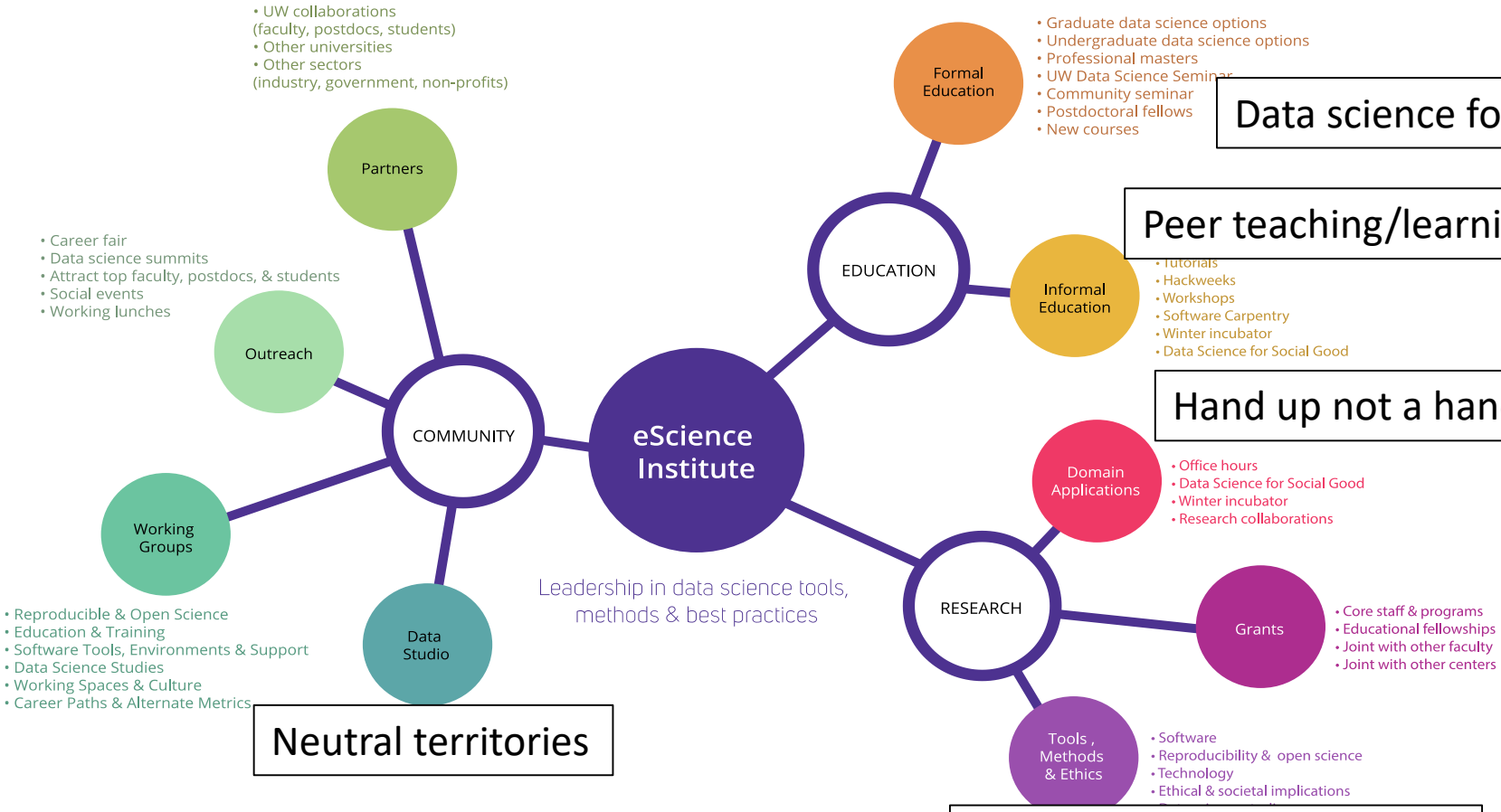
Posted Aug 26, 2016 by Devin Coldewey, Contributor

Next Story

Data Science @UW in 2014

Inaugural All-Campus Poster Session:
137 posters from 30+ departments




• UW collaborations
(faculty, postdocs, students)
• Other universities
• Other sectors
(industry, government, non-profits)
Partners
• Career fair
• Data science summits
• Attract top faculty, postdocs, & students
• Social events
• Working lunches
Outreach
COMMUNITY
Working Groups
Data Studio
• Reproducible & Open Science
• Education & Training
• Software Tools, Environments & Support
• Data Science Studies
• Working Spaces & Culture
• Career Paths & Alternate Metrics
Neutral territories
eScience Institute
Leadership in data science tools, methods & best practices
EDUCATION
Formal Education
• Graduate data science options
• Undergraduate data science options
• Professional masters
• UW Data Science Seminar
• Community seminar
• Postdoctoral fellows
• New courses
Data science for All
Informal Education
• Tutorials
• Hackweeks
• Workshops
• Software Carpentry
• Winter incubator
• Data Science for Social Good
Peer teaching/learning
RESEARCH
Domain Applications
• Office hours
• Data Science for Social Good
• Winter incubator
• Research collaborations
Hand up not a hand out
Grants
• Core staff & programs
• Educational fellowships
• Joint with other faculty
• Joint with other centers
Tools , Methods & Ethics
• Software
• Reproducibility & open science
• Technology
• Ethical & societal implications
Ethics in every conversation

Data Science Institutes (in partnership with Libraries) as Catalysts and Leaders

- Lead the advancement in data science methodologies
- Lead the utilization of these methodologies in discovery
- Lead the creation of environments where data science can flourish

The “Switzerland” of data science on campus

UNIVERSITY of WASHINGTON
eScience Institute
ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS

Thank You!

UNIVERSITY of WASHINGTON eScience Institute
ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS

Questions?

Micaela Parker & Sarah Stone

exec-director@escience.washington.edu

GORDON AND BETTY MOORE FOUNDATION

ALFRED P. SLOAN FOUNDATION 1934

NSF

Washington Research FOUNDATION

Links

“Creating Institutional Change in Data Science” Chronicles of Higher Ed, Mar 2018 (http://msdse.org/creating_institutional_change.html)

“Hack Weeks as a model for Data Science Education and Collaboration” (arXiv:1711.00028)

eScience Winter Incubator:
http://escience.washington.edu/get-involved/incubator-programs/overview/

UW Data Science for Social Good:
http://escience.washington.edu/dssg/